

**GILBERTO SANTOS DE OLIVEIRA**

**CARACTERIZAÇÃO DE *DISCRETE TYPING UNITS* (DTUS)  
UTILIZANDO PROTEÔMICA E FERRAMENTAS DE BIOINFORMÁTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

São Paulo  
2018

**GILBERTO SANTOS DE OLIVEIRA**

Caracterização de *Discrete Typing Units* (DTUs) utilizando proteômica e ferramentas de bioinformática

Dissertação apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

Área de concentração: Biologia da Relação Patógeno-Hospedeiro

Orientador: Prof. Dr. Giuseppe Palmisano

VERSÃO CORRIGIDA

São Paulo  
2018

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
Serviço de Biblioteca e informação Biomédica  
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Oliveira, Gilberto Santos de  
Caracterização de Discrete Typing Units (DTUs)  
utilizando proteômica e ferramentas de  
bioinformática / Gilberto Santos de Oliveira;  
orientador Giuseppe Palmisano. -- São Paulo, 2018.  
167 p.

Dissertação (Mestrado) ) -- Universidade de São  
Paulo, Instituto de Ciências Biomédicas.

1. Trypanosoma cruzi. 2. Discrete Typing Units.  
3. Espectrometria de massas. 4. Comparação  
espectral. 5. Método de tipificação. I. Palmisano,  
Giuseppe, orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

---

Candidato(a): Gilberto Santos de Oliveira

Título da Dissertação/Tese: Caracterização de Discrete Typing Units (DTUs) utilizando proteômica e ferramentas de bioinformática.

Orientador: Giuseppe Palmisano

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado/Tese de Doutorado, em sessão pública realizada a **22/03/2018**, considerou o(a) candidato(a):

**Aprovado(a)**

**Reprovado(a)**

Examinador(a):                      Assinatura: .....  
Nome: .....  
Instituição: .....

Examinador(a):                      Assinatura: .....  
Nome: .....  
Instituição: .....

Examinador(a):                      Assinatura: .....  
Nome: .....  
Instituição: .....

Presidente:                              Assinatura: .....  
Nome: .....  
Instituição: .....



Cidade Universitária "Armando de Salles Oliveira", Butantã, São Paulo, SP - Av. Professor Lineu Prestes, 2415 - ICB III - 05508 000  
Comissão de Ética em Pesquisa - Telefone (11) 3091-7733 - e-mail: cep@icb.usp.br

## CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB nº **779/2015** referente ao projeto intitulado: "*Caracterização de discrete typing units (DTUs) utilizando proteômica e ferramentas de bioinformática*" sob a responsabilidade de Gilberto Santos de Oliveira e orientação do(a) Prof.(a) Dr.(a) **Giuseppe Palmisano**, do Departamento de Parasitologia, foi analisado pela **CEUA** - Comissão de Ética no Uso de Animais e pela **CEPSH** - Comissão de Ética em Pesquisa com Seres Humanos, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da Lei nº 11.794, de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº 466 de 2012.

São Paulo, 25 de novembro de 2015

Prof. Dr. **Anderson de Sá Nunes**  
Coordenador CEUA ICB/USP

Prof. Dr. **Paolo M. A. Zanotto**  
Coordenador CEPSH ICB/USP

## AGRADECIMENTO

Primeiramente agradeço a minha família pelo apoio, especialmente minha mãe e minhas avós. A minha esposa Nathalia por sempre me apoiar e me ajudar a chegar até aqui, sempre me incentivando.

Agradeço também ao Prof. Dr. Giuseppe Palmisano por me dar a oportunidade de trabalhar com ele. Além do mais, conheci esse mundo maravilhoso da Proteômica e suas possibilidades por causa dele, sem falar na paciência de ensinar.

A Rebeca, que me ensinou muito e pelas nossas conversas descontraídas, principalmente no almoço, ótima pessoa para se trabalhar e conviver.

A Livia por me ajudar com os meus experimentos mesmo estando longe. Também a Claudia por me ajudar nas dúvidas que tenho no laboratório.

Aos meus colegas do lab, Joyce por termos conversas interessantes e também por eu sempre enchê-lá o “saco”. A Victória pelas nossas boas conversas até o bandeijão. Ao Daniel pela sua ajuda no lab e pelas nossas conversas descontraídas. A Ana por suas dicas, principalmente em relação a designer, não entendo nada disso (hahaha). E quero agradecer ao nosso mais novo integrante, Simon, mal chegou no lab e processamos 50 amostras (hahaha), obrigado pelo companheirismo.

Agradecimentos também ao Laboratório de Taxonomia e Filogenia de Tripanossomatídeos, principalmente a Profa. Dr. Marta Maria Geraldete Teixeira por ceder as cepas de *T. cruzi*, possibilitando este trabalho. E também a Marta Campaner por me ajudar nos cultivos dos parasitos.

Gostaria de agradecer também ao Prof. Dr. Martin R. Larsen por ceder o laboratório e a expertise.

E quero agradecer a todas as pessoas que me ajudaram até aqui. Obrigado a todos.

### **AGRADECIMENTO CNPQ**

Quero agradecer ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) (Processo 163579/2015-6), sem seu apoio nós ficaríamos impossibilitados de prosseguir com os estudos.

“A imaginação é mais importante  
que o conhecimento.”

- Albert Einstein



## RESUMO

OLIVEIRA GS, Caracterização de Discrete Typing Units (DTUs) utilizando proteômica e ferramentas de bioinformática. Dissertação (Mestrado em Parasitologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2017.

A doença de Chagas é uma das doenças negligenciadas mais importantes com um número estimado de 12 milhões de indivíduos infectados, a maioria vivendo na América Central e do Sul. O parasito protozoário *Trypanosoma cruzi* (*T. cruzi*) é o agente etiológico da doença de Chagas. O *T. cruzi* é geneticamente diversificado. Uma nova nomenclatura foi adotada para classificar as cepas de *T. cruzi* (TcI-TcVI e Tcbat), denominados *Discrete Typing Units* (DTUs), com base em suas características bioquímicas, imunológicas e fenotípicas. As DTUs de *T. cruzi* foram correlacionadas a diversos desfechos clínicos, destacando a importância de testes epidemiológicos e moleculares específicos. Apesar do desenvolvimento de métodos de caracterização de *T. cruzi* baseados em assinaturas genéticas, cada método apresenta vantagens e desvantagens. O nosso trabalho mostra a aplicação da espectrometria de massas para o ensaio de caracterização de cepas de *Trypanosoma cruzi* usando bibliotecas espectrais MS<sup>2</sup> de peptídeos (*Trypanosoma cruzi* Strain Typing Assay using MS<sup>2</sup>, Tc-STAMS<sup>2</sup>). A novidade do método é o uso de espectros de fragmentação de peptídeos como impressões digitais específicas para classificar e identificar as DTUs. Inicialmente, uma biblioteca de espectros é gerada a partir das DTUs de *T. cruzi*. A biblioteca é subsequentemente desafiada usando espectros MS/MS das DTUs desconhecidas e confiavelmente atribuí a uma DTU específica em uma abordagem automática e computacional. Utilizando o método Tc-STAMS<sup>2</sup> foram testadas diferentes variáveis como preparação e tipo de amostra, configuração do instrumento e plataforma de identificação, proporcionando alta confiança e robustez na caracterização das DTUs de *T. cruzi*. O método Tc-STAMS<sup>2</sup> representa uma estratégia complementar aos métodos atuais de genotipagem de *T. cruzi* baseados em DNA. Além disso, o método permite a identificação de características específicas das DTUs que poderiam estar relacionadas à biologia de *T. cruzi*.

**Palavras-chave:** *Trypanosoma cruzi*. *Discrete Typing Units* (DTUs). Espectrometria de massas. Comparação espectral. Métodos de tipificação.

## ABSTRACT

OLIVEIRA GS, Characterization of Discrete Typing Units (DTUs) using proteomics and bioinformatics tools. 2017. Masters thesis (Parasitology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2017.

Chagas disease is one of the most important neglected diseases with an estimated number of 12 million infected individuals, the majority living in Central and South America. The *Trypanosoma cruzi* (*T. cruzi*) protozoan parasite is the etiological agent of Chagas' disease. *T. cruzi* is highly genetically diverse and a new nomenclature separated each strain into seven different genetic groups (TcI-TcVI and Tcbat), named Discrete Typing Units (DTUs) based on their biochemical, immunological and phenotypical characteristics. *T. cruzi* DTUs have been correlated to diverse clinical outcomes highlighting the importance of molecular epidemiological screens. Despite the development of *T. cruzi* typing methods based on genetic signatures, each method presents its own advantages and challenges. The work presented here shows the application of mass spectrometry for *Trypanosoma cruzi* Strain Typing Assay using MS<sup>2</sup> peptide spectral libraries (Tc-STAMS<sup>2</sup>). The novelty of the method is based on the use of peptide fragmentation spectra as strain-specific fingerprints to classify and identify DTUs. Initially, a spectra library is generated from characterized *T. cruzi* strains. The library is subsequently inspected using MS/MS spectra from unknown strains and confidently assigned to a specific strain in an automated and computationally driven approach. The Tc-STAMS<sup>2</sup> method was employed to test several variables such as sample type and preparation, instrument setup and identification platform. Tc-STAMS<sup>2</sup> provided high confidence and robustness in *T. cruzi* strain typing. The Tc-STAMS<sup>2</sup> method represents a proof-of-concept of a complementary strategy to the current DNA-based *T. cruzi* genotyping methods. Moreover, the method allows the identification of strain-specific features that could be related to the biology of *T. cruzi* strains and their clinical outcomes.

**Keywords:** *Trypanosoma cruzi*. Discrete Typing Units (DTUs). Mass spectrometry. Spectral matching. Strain typing methods.

## LISTAS DE FIGURAS

**Figura 1** - Distribuição geográfica da Doença de Chagas ao redor do mundo.

**Figura 2** - Ciclo do *Trypanosoma cruzi*.

**Figura 3** - Distribuição epidemiologia das DTUs em humanos na América do Sul.

**Figura 4** - Representação da maquinaria de transcrição e processamento de mRNAs em tripanossomatídeos.

**Figura 5** - Vias de entrada do parasito.

**Figura 6** - Visão geral da proteômica baseada em espectrometria de massas *bottom-up*.

**Figura 7** - Componentes do espectrômetro de massas.

**Figura 8** - Representação dos íons gerados nos diferentes métodos de fragmentação.

**Figura 9** - Visão geral do funcionamento do LTQ-Orbitrap Velos em modalidade aquisição dependentes de dados (DDA).

**Figura 10** - Identificações dos peptídeos

**Figura 11** - Fluxograma para a construção de uma biblioteca espectral.

**Figura 12** - Tc-STAMS<sup>2</sup> fluxograma.

**Figura 13** - Perfil cromatográfico LC-MS dos peptídeos pertencentes as quatro réplicas da cepa *Sylvio X10 cl1* (DTU-I).

**Figura 14** - Perfil cromatográfico LC-MS dos peptídeos pertencentes das seis DTUs.

**Figura 15** - Matrix de validação do método Tc-STAMS<sup>2</sup>.

**Figura 16** - Perfil cromatográfico LC-MS dos peptídeos pertencentes às cepas CL Brener (DTU-VI) e *CL14* do *T. cruzi*.

**Figura 17** - Teste do Tc-STAMS<sup>2</sup> quanto a sua robustez para variações técnicas e experimentais.

**Figura 18** - Perfil cromatográfico LC-MS dos peptídeos pertencentes à DTU-I adquirido no Grupo PR em Odense na Dinamarca e na *facility* de espectrometria de massas do CEFAP na USP, São Paulo, Brasil.

**Figura 19** - Identificação dos peptídeos utilizando os softwares MaxQuant, TPP e Proteome Discoverer<sup>TM</sup>.

**Figura 20** - *Heat map* das A) Proteínas e B) Peptídeos diferencialmente regulados.

**Figura 21** - Análise de componentes principais (PCA) das proteínas reguladas das seis DTUs somadas a cepa *CL14*.

**Figura 22** - *Multi Scatter Plot* obtido comparando as proteínas quantificadas das quatro réplicas das seis DTUs utilizando cada uma das proteínas quantificadas.

**Figura 23** - *Multi Scatter Plot* obtido comparando os peptídeos quantificados das seis DTUs.

**Figura 24** - Rede de interações proteína-proteína utilizando 306 proteínas construída utilizando o software String.

## LISTAS DE TABELAS

**Tabela 1** - Representação das cepas e sua correspondência nas DTUs.

**Tabela 2** – Lista de cepas utilizadas para o desenvolvimento e validação do método Tc-STAMS<sup>2</sup>

**Tabela 3** - Identificações das cepas de *T. cruzi* baseados nas buscas de similaridade espectral.

**Tabela 4** - Comparação das correspondências espectrais MS/MS usando diferentes fases de crescimento de *Sylvio X10/1* (DTU-I).

**Tabela 5** - Todas as identificações das DTUs e *T. cruzi-like*.

**Tabela 6** - Correspondência espectral entre as diferentes DTUs utilizando o DiagnoProt.

**Tabelas suplementares apresentadas no CD em anexo:**

**Tabela suplementar S1, S2 e S3** - Números de IDs das proteínas usando as plataformas do MaxQuant, TPP e Proteome Discoverer.

**Tabela suplementar S4** - Números de PSMs e espectros MS/MS não anotados.

**Tabela Suplementar S5** - Peptídeos e proteínas reguladas entre as seis DTUs.

**Tabela suplementar S6** - Tabela com os dados da rede de interação proteína-proteína.

## LISTAS DE ABREVIações

18S rRNA - Subunidade 18S do RNA ribossômico (*18S ribosomal RNA*)

24S  $\alpha$ -rRNA - Subunidade 24S  $\alpha$  do RNA ribossômico (*24S  $\alpha$ -subunit rRNA genes*)

2DE – Gel de Eletroforese Bidimensional (*Two-Dimensional Gel electrophoresis*)

2DE-DIGE - Gel de Eletroforese Bidimensional Diferencial Fluorescente (*Fluorescent Difference Gel Electrophoresis*)

ACN - Acetonitrila

ANOVA - Análise de Variância

BSF1 - Forma Sanguínea (*Bloodstream-form*) (Nessa tese associada ao *T. vivax*)

CC - Corrente Contínua

CEFAP - Centro de Facilidades para a Pesquisa

CID - Dissociação Induzida por Colisão (*Collision Induced Dissociation*)

DC - Doença de Chagas

DDA - Aquisição Dependente de Dados (*Data-Dependent Acquisition*)

DIT MALDI-TOF MS - Identificação Direta de Tripanossomatídeos por MALDI-TOF MS (*Direct identification of trypanosomatids of Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry*)

DNA - Ácido Desoxirribonucleico (*Deoxyribonucleic Acid*)

DTT - Ditioneitol (*Dithiothreitol*)

DTUs - *Discrete Typing Units*

*E. coli* - *Escherichia coli*

ECD - Dissociação por Captura de Elétrons (*Electron Capture Dissociation*)

ELISA - Ensaio de Imunoabsorção (*Enzyme-Linked Immunosorbent Assay*)

EP1 - Forma Epimastigota (Nessa tese associada ao *T. vivax*).

ESI - Ionização por Electrospray (*Electrospray Ionization*)

ETD - Dissociação por Transferência de Elétrons (*Electron Transfer Dissociation*)

eV - Elétron-Volt

FDR - Taxa de Falsos Positivos (*False Discovery Rates*)

FT-ICR - Ressonância Ciclotrônica de Íons por Transformada de Fourier (*Fourier-Transform Ion Cyclotron Resonance*)

FWHM - Largura a Meia Altura (*Full Width at Half Maximum*)

gp63 - Glicoproteína 63

HAI - Ensaio de Hemaglutinação Indireta (*Indirect Hemagglutination Assay*)

HCD - Dissociação Induzida por Alta Energia de Colisão (*High Energy Collisional Dissociation*)

HSP - Proteínas de Choque Térmico (*Heat-Shock Proteins*)

ICAT - Marcador de Afinidade Enriquecido Isotopicamente (*Isotope-Coded Affinity Tags*)

IFI - Teste de Imunofluorescência (*Immunofluorescence Test*)

IgG - Imunoglobulina G

IgM - Imunoglobulina M

IT - Armadilha de Íons (*Ion Trap*)

iTRAQ - Marcador Isobárico para Quantificação Relativa e Absoluta (*Isobaric Tags for Relative and Absolute Quantification*)

kDNA - Cinetoplasto DNA (*kinetoplast DNA*)

keV - Kilo Elétron-Volt

LC - Cromatografia Líquida (*Liquid Chromatography*)

LC-MS - Cromatografia Líquida acoplada a Espectrometria de Massas (*Liquid Chromatography Mass Spectrometry*)

LFQ - Quantificação Livre de Marcador (*Label-Free Quantification*)

LIT - Armadilha de Íons Linear (*Linear Ion Trap*)

LIT Medium – Meio Infusão de Fígado e Triptose (*Liver Infusion Tryptose Medium*)

$m/z$  - Massa/Carga (*mass/charge*)

m7G cap - 7-Metilguanosina

MALDI - Ionização por Dessorção a Laser Assistida por Matriz (*Matrix-Assisted Laser Desorption/Ionization*)

MALDI-TOF - Ionização por Dessorção a Laser Assistida por Matriz por Tempo de Vôo (*Matrix-Assisted Laser Desorption/Ionization Time-of-Flight*)

MASPs - Proteínas de Superfície Associadas a Mucinas (*Mucin-Associated Surface Proteins*)

Meta3 - Forma Metacíclica (Nessa tese associada ao *T. vivax*)

MLEE – Enzima de Eletroforese Multilocus (*Multilocus Enzyme Electrophoresis*)

mRNA - RNA mensageiro

MLST - Tipificação de Sequência por Multi-locus (*Multi-Locus Sequence Typing*)

MS/MS - Espectrometria de Massas Sequencial (*Tandem-Mass Spectrometry*)

Mtq-PCR -PCR em Tempo Real Multiplex (*Multiplex Real-Time PCR*)

NIST - Instituto Nacional de Padrões e Tecnologia (*National Institute of Standards and Technology*)

N-Terminal - Amino-Terminal

nLC-MS/MS - nano-LC-MS/MS (*nano-Liquid Chromatography Tandem-Mass Spectrometry*)

PARP - Proteína Repetitiva Ácida Procíclica (*Procylic Acidic Repetitive Protein*)

PBS - Salina Tamponada com Fosfato (*Phosphate-Buffered Saline*)

PCA - Análise de Componentes Principais (*Principal Component Analysis*)

PCR - Reação em Cadeia da Polimerase (*Polymerase Chain Reaction*)

PMF - Impressão Digital da Massa do Peptídeo (*Peptide Mass Fingerprint*)

Poli-A - Cauda Poli-A

pré-mRNA - Pré-RNA mensageiro (*Pre-messenger Ribonucleic Acid*)

PRIDE - Proteomics IDentifications

PTMs - Modificações Pós-Traducionais (*Post-Translational Modifications*)

Q-Q-LIT - Quadropolo-Armadilha de ions

RAPD - DNA Polimórfico Amplificado Aleatório (*Random Amplified Polymorphic DNA*)

RF - Rádio Frequência

RFLP - Polimorfismo do Comprimento do Fragmento de Restrição (*Restriction Fragment Length Polymorphism*)

HSPs - Proteínas de Choque Térmico (*Heat Shock Proteins*)

RNA - Ácido Ribonucleico (*Ribonucleic Acid*)

SC - Contagem Espectral (*Spectral Counting*)

SCX - Troca Catiônica Forte (*Strong Cation Exchange*)

SDSS - Similaridade de Conjuntos de Dados Espectrais (*Spectral Dataset Similarity*)

SILAC - Rotulagem Isotópica Estável por Aminoácidos em Cultura de Células (*Stable Isotope Labeling by Amino Acids in Cell Culture*)

S<sub>p</sub> - Score Preliminar (*Preliminary Score*)

SRM - Monitoração de Reação Seleccionada (*Selected Reaction Monitoring*)

*T. cruzi* - *Trypanosoma cruzi*



## SUMÁRIO

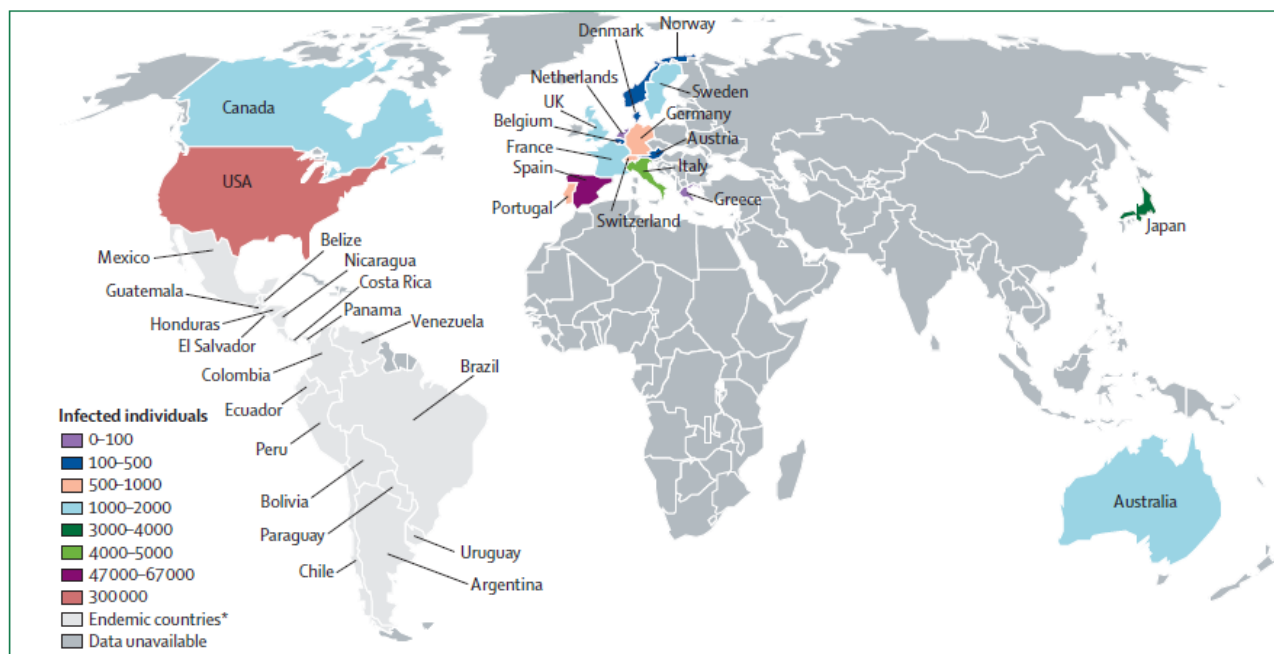
1 INTRODUÇÃO .....	19
1.1 <i>Trypanosoma cruzi</i> .....	19
1.2 Prevalência .....	20
1.3 Mortalidade.....	20
1.4 Gênero <i>Trypanosoma</i> .....	20
1.5 Ciclo de Vida do <i>Trypanosoma cruzi</i> .....	21
1.6 Discrete Typing Units.....	23
1.7 Genômica e expressão gênica do <i>T. cruzi</i> .....	26
1.8 Doença de Chagas.....	28
1.8.1 Fase Aguda .....	28
1.8.2 Fase Crônica .....	29
1.8.3 Tratamento .....	29
1.9 Diagnóstico da Doença de Chagas .....	30
1.10 Proteômica Baseada em Espectrometria de Massas .....	30
1.10.1 Avanços Tecnológicos na Proteômica Baseada em Espectrometria de Massas .....	32
1.10.2 Fragmentação de Peptídeos.....	33
1.10.3 Análise de Dados Proteômicos .....	36
1.11 Biblioteca Espectral .....	37
1.12 Proteômica no <i>Trypanosoma cruzi</i> .....	41
2 OBJETIVOS .....	45
2.1 Objetivo Geral.....	45
2.2 Objetivos Específicos .....	45
3 MATERIAL E MÉTODOS .....	46
3.1 Cultura Celular de <i>Trypanosoma cruzi</i> .....	46
3.2 Influência das Condições de Cultivo do <i>Trypanosoma cruzi</i> .....	47
3.3 Preparação das Amostras e Análise nLC-MS/MS.....	47
3.3.1 Preparação da Amostra para as Condições Ácidas e Básicas .....	47
3.3.2 Análises Nano LC-MS/MS.....	48
3.3.4 Quantidades de Amostra, Gradientes Cromatográficos e Tipos de Fragmentação MS/MS Utilizados no Teste de Conceito. ....	48
3.4 Bioinformática e Análises Estatísticas.....	49
3.4.1 Geração da Biblioteca Espectral MS/MS e Correspondência Espectral .....	49
3.4.2 Busca em Banco de Dados.....	50
3.4.3 Análise de Redes Interações Proteína-Proteína .....	51

3.4.4 Análises Estatísticas .....	51
4 RESULTADOS .....	52
4.1 A Estratégia Tc-STAMS <sup>2</sup> Permitiu a Discriminação das DTUs.....	52
4.2 Análise de Clustering Utilizando Identificações de Peptídeos e Proteínas .....	62
4.3 Análise de Rede Interação Proteína-Proteína .....	70
5 DISCUSSÃO .....	72
6 CONCLUSÃO .....	77
REFERÊNCIAS* .....	78
APÊNDICES .....	91
APÊNDICE A .....	91
APÊNDICE B .....	125
APÊNDICE C .....	164

# 1 INTRODUÇÃO

## 1.1 Trypanosoma cruzi

A Doença de Chagas (DC), também conhecida como Tripanossomíase americana, foi descoberta e descrita por Carlos Ribeiro Justiliano Chagas em 1909 (1). De acordo com a Organização Mundial de Saúde (*World Health Organization*, WHO), a doença afeta por volta de 8 milhões de pessoas, principalmente na América Latina, e é causada pelo protozoário *Trypanosoma cruzi* (*T. cruzi*) (2). Estima-se que 10 mil pessoas morram todos os anos devido às manifestações clínicas da Doença de Chagas e que cerca de 25 milhões de pessoas corram risco de adquirir a doença (2) (**Figura 1**). A transmissão da DC pode ocorrer por animais silvestres (ex. gambá, tatus) e domésticos infectados (ex. Cachorro e gato), por transfusões sanguíneas e de forma congênita (3, 4). Transplantes de órgãos e ingestão de bebidas contaminadas representam outras formas de infecção (4). Contudo, a principal forma de transmissão ocorre pela picada do inseto contaminado, os insetos hematófagos da subfamília *Triatominae*, principalmente o *Triatoma infestans*, *Triatoma dimidiata* e *Rhodnius prolixus*, conhecidos vulgarmente como "barbeiros" (2).



**Figura 1** – Distribuição geográfica da Doença de Chagas ao redor do mundo (4).

No Brasil, um país de proporções continentais que vem passando por rápidas transformações demográficas, sociais e ambientais, onde as desigualdades socioeconômicas e regionais continuam muito evidentes, doenças negligenciadas atingem grande parte da população (5). Estima-se que no Brasil o número de pessoas com DC varie entre 1,9 a 4,6 milhões (6-8). Devido ao número elevado de doentes, é um grande desafio para os próximos anos estabelecer e sustentar planos consistentes para o diagnóstico e tratamento de dos pacientes (9).

## 1.2 Prevalência

Em relação à morbimortalidade da DC no Brasil, há poucos estudos sistemáticos de base populacional, dificultando sua estimativa (8). Por volta de 1950 a DC era reconhecida como uma endemia rural, porém, com o processo de industrialização do Brasil e aumento da migração da população rural para áreas urbanas, a doença foi tomando caráter urbano (8). Apenas a partir de 2014, quando foi publicado o primeiro estudo utilizando dados de publicações de 1980 a 2012, chegou-se a prevalência da DC no Brasil: 4,2% variando de 4,4% na década de 1980 a 2,4% após o ano 2000 (6). Tendo como base esse e outros estudos, a variação fica por volta de 1,0% a 2,4% da população (6-8).

## 1.3 Mortalidade

A mortalidade da DC no Brasil persiste em níveis elevados (6, 10-13). Em um estudo realizado utilizando dados de 1999 a 2007 foram identificados 53.924 óbitos relacionados à DC, sendo 44.537 tendo a DC como causa básica e 9.387 como causa associada (12). Adicionalmente, entre os períodos de 2000 a 2010, notou-se que as mortes causadas pela DC ocorreram preferencialmente em pessoas do sexo masculino acima de 60 anos (85,9%) (14). Em um período de 14 anos, a letalidade anual média foi de 2,7% (37,9). Entre os anos de 2005 e 2013 houve um aumento da letalidade, que passou para 20,0%, período que coincide com o surto de transmissão oral ocorrido em Santa Catarina no ano de 2005 (15). A letalidade continuou elevada em 2006 (5,9%), porém diminuindo constantemente ao passar dos anos (15).

## 1.4 Gênero *Trypanosoma*

O *Trypanosoma cruzi* pertence ao filo Euglenozoa, do qual fazem parte a ordem dos Kinetoplastídeos Euglenoidea e Diplonemea (16). A característica da ordem dos Kinetoplastídeos é sua mitocôndria única, o kDNA (17). O kDNA contém em sua composição moléculas dupla-fita circulares denominadas maxicírculos e minicírculos que formam uma rede, sendo responsável pela

codificação de proteínas ou RNAs mitocondriais que são responsáveis pela edição dos genes transcritos (18).

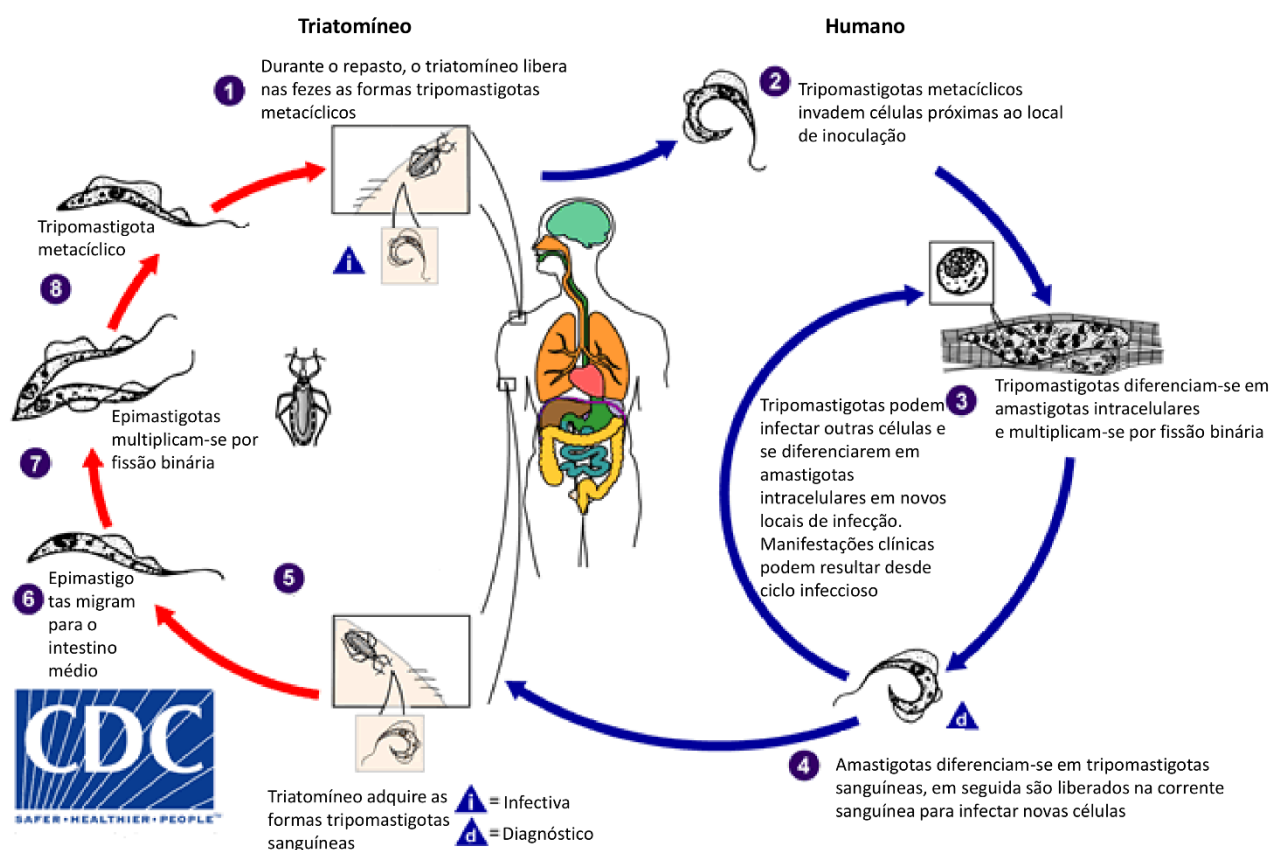
A ordem dos Kinetoplastídeos se divide em dois subgrupos, Prokinetoplastina, que engloba o subgrupo Prokinetoplastida e Metakinetoplastina, que possui quatro subordens: os biflagelados Bodonídeos (Neobodonida, Parabodonida e Eubodonida) e a ordem Trypanosomatida, dentro da qual se encontra a família dos uniflagelados Trypanosomatidae (17, 19, 20).

Os parasitos da família Trypanosomatidae podem ser encontrados nos mais variados ambientes, parasitando todas as classes de vertebrados, invertebrados e plantas, (21-25). A família Trypanosomatidae é composta por 14 gêneros, sendo divididos em monoxênicos e heteroxênicos. Os monoxênicos, que utilizam em seus ciclos de vida hospedeiros invertebrados como os artrópodes hematófagos da ordem Diptera (moscas e mosquitos) e Hemiptera (triatomíneos), compreende os gêneros *Paratrypanosoma*, *Blechnomonas*, *Crithidia*, *Blastocrithidia*, *Wallaceina*, *Leptomonas*, *Herpetomonas*, *Sergeia*, *Strigomonas* e *Angomonas*. Já os heteroxênicos, cujo ciclo de vida envolve hospedeiros invertebrados e vertebrados (mamíferos, anfíbios, répteis, peixes e aves) ou invertebrados e vegetais, compreende os gêneros *Phytomonas*, *Endotrypanum*, *Leishmania* e *Trypanosoma* (21-27).

### 1.5 Ciclo de Vida do *Trypanosoma cruzi*

Como descrito por Carlos Chagas em 1909 (1), o *T. cruzi* apresenta um ciclo de vida complexo que ocorre em dois hospedeiros: insetos triatomíneos hematófagos e mamíferos. Durante seu ciclo de vida o *T. cruzi* apresenta três formas distintas: amastigota, epimastigota e tripomastigota. O triatomíneo durante o repasto sanguíneo no mamífero adquire as formas tripomastigotas sanguíneas (**Figura 2**). Os tripomastigotas sanguíneos migram para o intestino médio do inseto e se diferenciam em epimastigota, que são formas replicativas. Os parasitos multiplicam-se por divisão binária e, em seguida, migram para a porção posterior do intestino. Na porção final, as formas epimastigotas ligam-se à parede do intestino posterior (ampola retal) antes de se diferenciarem em tripomastigotas metacíclicos (28, 29). A metaciclogênese parece ser desencadeada pela interação hidrofóbica entre o flagelo e o substrato ao qual ele se anexa (28). Após se diferenciarem em tripomastigotas metacíclicos, estes se separam da parede do intestino e são excretados junto com as fezes (30). As formas tripomastigotas metacíclicas são capazes de infectarem uma ampla gama de células e tecidos de mamíferos como macrófagos, células epiteliais, fibroblastos e musculatura lisa e estriada (31).

Durante o repasto sanguíneo, próximo ao local da picada, o inseto libera nas fezes as formas tripomastigotas metacíclicas, que entram no hospedeiro através do ferimento causado pela picada ou por ferimentos causados pelo ato de coçar, realizado pelo próprio mamífero. No hospedeiro (**Figura 2**), os tripomastigotas metacíclicos são internalizados por fagocitose por macrófagos, ou por outras células do hospedeiro por endocitose (31). Essa internalização dá origem ao chamado vacúolo parasitóforo, local onde o parasito se encontra. Os vacúolos parasitóforos são formados por fusões de lisossomos com a membrana plasmática (32). O interior ácido (pH abaixo de 6.0) do vacúolo parasitóforo parece favorecer o escape dos parasitos por ativar a molécula *porin-like* Tc-Tox que apresenta homologia com a perforina, proteína que é responsável por criar poros na membrana plasmática (33). Após formas tripomastigotas metacíclicas escaparem do vacúolo parasitóforo e passarem para citosol, estas começam a se diferenciar em amastigotas intracelulares. Os amastigotas se multiplicam por divisão binária, se diferenciam em tripomastigotas sanguíneas e em seguida, são liberados na corrente sanguínea para infectar novas células e refazer o ciclo (**Figura 2**) (4).



**Figura 2** – Ciclo *Trypanosoma cruzi*, 1- Dentro do mamífero, que no exemplo é o ser humano. Os tripomastigotas metacíclicos invadem células próximas ao local de invasão, se diferenciam em amastigotas. 2 – Amastigotas se diferenciam por divisão binária. 3 – Os tripomastigotas sanguíneos saem da célula e caem na corrente sanguínea. 4 – Os tripomastigotas invadem novas células e repetem o ciclo, ou 5 - são ingeridos por outro triatomíneo durante o repasto sanguíneo. 6 - Os tripomastigotas ingeridos se diferenciam em epimastigotas no intestino médio do inseto. 7 – Os epimastigotas se multiplicam por divisão binária. 8 – Diferenciam-se em tripomastigotas metacíclicos na porção fina do intestino. Fonte: modificado de: Centros de Controle e Prevenção de Doenças (CDC) (34).

## 1.6 Discrete Typing Units

O *T. cruzi* é geneticamente e morfologicamente diverso. Com o objetivo de uniformizar a nomenclatura e a comunicação entre os cientistas, as cepas do *T. cruzi* foram inicialmente divididas em seis grupos (*T. cruzi* I-VI) de *Discrete Typing Units* (DTUs) (**Tabela 1 e Figura 3**) (35, 36), uma nova cepa associada a morcegos foi caracterizada (Tc-Bat/DTU-VI) (37, 38). Cada grupo representa um conjunto de isolados que são geneticamente semelhantes e que podem ser identificados por marcadores moleculares ou imunológicos comuns (**Tabela 1**) (35).

**Tabela 1** - Representação das cepas e sua correspondência nas DTUs. Alguns exemplos são reportados.

Cepa	DTU	País	Hospedeiro/Vetor
<i>Sylvio X10 cl1</i>	I	Pará, Brasil	<i>Homo sapiens</i>
Y	II	São Paulo, Brasil	<i>Homo sapiens</i>
3869	III	Amazonas, Brasil	<i>Homo sapiens</i>
<i>CanIII cl1</i>	IV	Pará, Brasil	<i>Homo sapiens</i>
<i>MN cl2</i>	V	Region IV, Chile	<i>Homo sapiens</i>
<i>CL Brener</i>	VI	Rio Grande do Sul, Brasil	<i>Triatoma infestans</i>

Nota - Tabela com as diferentes DTUs e sua distribuições. Adaptado de Zingales et al (35) .

Anteriormente à atribuição das DTUs, os isolados de *T. cruzi* eram classificados de acordo com alguns critérios, como por exemplo, sua morfologia. Outros critérios como virulência, suscetibilidade a fármacos, patogenicidade e perfil eletroforético de isoenzimas (39-43) também foram utilizados. Além disso, estudos utilizando o perfil eletroforético de isoenzimas permitiu a separação de três grupos principais, nomeados como Zimodemas (44, 45). Posteriormente, Tibayrenc e Ayala com a utilização de maior número de isolados foram capazes de aumentar significativamente o número de zimodemas, que agora conta com 43 zimodemas (43). Estas evidências reforçaram a ideia que o *T. cruzi* é um organismo diplóide, que sua reprodução ocorre majoritariamente por divisão binária (reprodução assexuada) e que a reprodução por processo de hibridização (reprodução sexuada) é rara. Esses fatos mostram a estabilidade do organismo que ocorreu ao longo do tempo (45, 46).

A caracterização das cepas de *T. cruzi* é extremamente importante para entender as diferentes características epidemiológicas e patológicas, sua distribuição geográfica e suas manifestações clínicas (35). Devido a isso, várias metodologias de caracterização tem sido introduzidas para melhorar a genotipagem das cepas de *T. cruzi*. Em particular, as diversidades genéticas das populações de *T. cruzi* foram reconhecidas primeiramente por MLEE (*Multilocus Enzyme Electrophoresis*) (44, 47) (46), RAPD (*Random Amplified Polymorphic DNA*) (48) e posteriormente por análises de restrição do DNA (49), gel de eletroforese em campo pulsado (50) e DNA *fingerprint* (51), demonstrando que existe uma infinidade de métodos para tipificação.

As metodologias mencionadas anteriormente demonstraram que algumas DTUs consistem de linhagens híbridas. Como revisado por Zingales et al (36), as linhagens III e IV podem ter se formado a partir de um ancestral em comum entre I e II. Adicionalmente, as linhagens V e VI podem



ter ocorrido pela combinação entre II e III (52-55). Portanto, conclui-se que, embora raros, trocas genéticas entre as cepas podem acontecer (56).

Quando o parasito infecta um hospedeiro, o mesmo pode sofrer uma seleção clonal (57), com isso, os parasitos isolados de insetos e humanos podem apresentar genótipos diferentes entre si (57). A heterogeneidade genética do *T. cruzi* torna difícil o entendimento epidemiológico e patológico da DC. Ainda está pouco elucidado até que ponto a infecção por diferentes variantes do *T. cruzi* influencia a severidade da doença. Além disso, tem-se demonstrado que outros fatores podem estar associados, como fatores geográficos, a linhagem do parasito, aspectos clínicos da doença e variação genética entre o parasito e o hospedeiro (56).



**Figura 3** – Distribuição epidemiologia das DTUs em humanos na América do Sul. Fonte: Adaptado de: Zingales et al; Lima L, Espinosa-Alvares O, Ortiz PA et al (38, 58).

Atualmente, os métodos utilizados para tipificação das linhagens de *T. cruzi* são baseados na caracterização dos genes do ribossomo, 24S  $\alpha$ -rRNA e 18S rRNA. Outras análises envolvem

amplificação do DNA genômico (59), porém, necessita padrões complexos eletroforéticos (60). Esses métodos são capazes de discriminar as linhagens, entretanto, as interpretações dos resultados podem levar ao erro, como por exemplo a interpretação das análises de diferenças entre bandas de géis de eletroforese (61). Outro método introduzido recentemente é a MLST (Tipificação de Sequências Multilocus) que tem a capacidade de discriminar genótipos únicos em DTUs, porém, a sua implementação não é simples para uma caracterização rápida em grandes coleções de isolados (59). Métodos simples de caracterização como o RFLP (Polimorfismo de Comprimento de Fragmento) e o MTq-PCR (PCR em Tempo Real Multiplex) têm sido aplicados em amostras de isolados clínicos demonstrando seus potenciais em aplicações na clínica (62).

### 1.7 Genômica e expressão gênica do *T. cruzi*

O *T. cruzi* teve seu primeiro genoma sequenciado em 2005 por El-Sayed et. al., utilizando a metodologia de WGS (*Whole-Genome Shotgun*), na qual o DNA é clivado aleatoriamente e os fragmentos são sequenciados pelo método de terminação de cadeia (63). A cepa utilizada para o estudo foi a cepa híbrida *CL Brener* (DTU-VI) (64). A montagem atual do genoma (ASM20906v1) tem tamanho de 89 Mb, com 23,696 genes codificando 19,607 proteínas. Adicionalmente, 50% genoma do *T. cruzi* contém sequências repetidas, consistindo em um grande número de genes de proteínas de superfície como as MASPs (*Mucin-Associated Surface Proteins*), transialidases (TS), glicoproteína gp63, retrotransposons e repetições subteloméricas (64). Além do mais, mais de 50% das proteínas são reportadas como *hypothetical* ou *uncharacterized*. Essa falta de anotação, prejudica a associação com funções específicas e mais estudos são necessários para identificar e validar a função dessas proteínas.

Em particular, pelo fato da cepa *CL Brener* (DTU-VI) ser uma hibridização entre duas divergentes DTUs, como as *Esmeraldo-like* (DTU-II) e *non-Esmeraldo-like* (DTU-III), e pelo seu conteúdo genômico repetitivo, a montagem fica prejudicada, afetando a primeira montagem do genoma do *T. cruzi* e deixando-o incompleto (64, 65). Com a ideia de obter um genoma mais completo, Frazén et al., (65) sequenciaram uma cepa não-híbrida, a cepa *Sylvio X10/1* (DTU-I). A cepa *Sylvio X10/1* contém poucas repetições, baixos níveis de heterozigose, genoma menor e também é uma referência em estudos *in vivo* e *in vitro* (65, 66).

Ao comparar as cepas *Sylvio X10/1* (DTU-I) e *CL Brener* (DTU-VI) Frazén et al., notaram que os genomas das cepas eram similares e que continham regiões conservadas intercaladas com regiões de sequências repetitivas. Apesar da boa cobertura do genoma, as regiões que continham repetições prejudicaram a montagem do genoma. Para contornar essa dificuldade, os autores

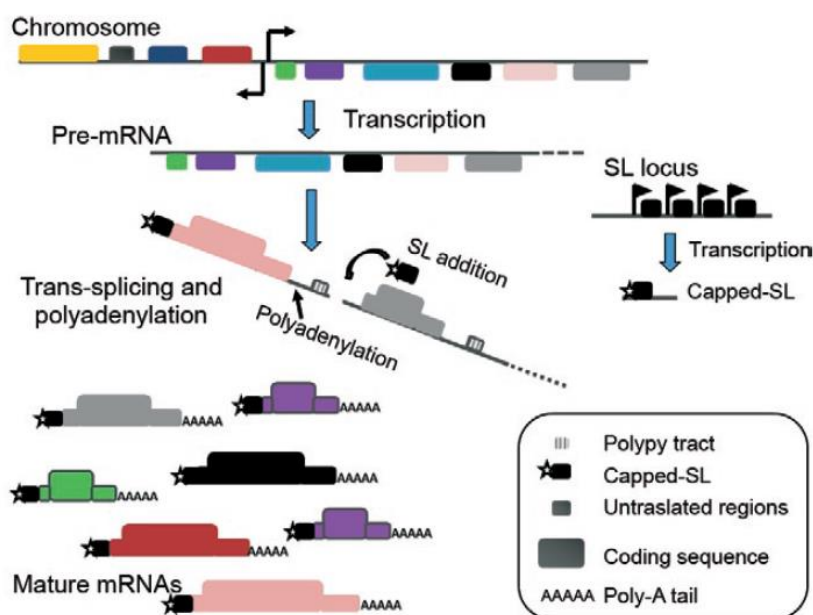
complementaram a montagem baseando-se na análise de leitura. Sendo assim, foi possível caracterizar os genes repetidos de ambos genomas. A comparação entre os representantes das DTU-I e DTU-VI mostraram que há uma grande diferença entre as duas cepas, principalmente em relação a proporção de sequências com homologia multigênicas. Adicionalmente, as análises dessas duas cepas deixaram clara a sua diferença, a cepa *CL Brener* possui um genoma de aproximadamente de 55 Mb, esse número está relacionado a expansão de genes relacionados a proteínas de superfície, enquanto que cepa *Sylvio X10/1* possui um genoma de aproximadamente de 44 Mb. Os autores notaram que o tamanho menor do genoma é característica da DTU-I (65).

Além das cepas *Sylvio X10/1* (DTU-I) e *CL Brener* (DTU-VI), outras cepas também foram sequenciadas tais como o *T. cruzi* JR CL4 (genBank assembly accession: GCA 000331405.1), *T. cruzi* Tula CL2 (genBank assembly accession GCA 000365225.1), *T. cruzi* Dm28c (67) e *T. cruzi marinkellei* cepa B7 (68). Muitas dessas cepas encontram-se atualmente com o genoma em “draft”.

Devido a mudanças de hospedeiros durante o ciclo biológico, o *T. cruzi* sofre alterações morfológicas e moleculares que são consequência da expressão gênica e das vias metabólicas (69). Essas alterações necessitam de um controle refinado e rápido da expressão gênica. Desde a transcrição do DNA até a tradução do RNA mensageiro, o parasito desenvolveu mecanismo diferentes de outros eucariotos, gerando processos únicos na regulação (70), tais como: a transcrição policistrônica (71-73), o processamento do pré-mRNA por *trans-splicing* e poliadenilação (74, 75) e um mecanismo para a reconstituição da fase de leitura aberta do transcrito, no qual a uridina é removida ou inserida por edição do pré-RNA mitocondrial (73, 76, 77). Adicionalmente, em alguns tripanossomatídeos, a RNA polimerase I é responsável pela transcrição de alguns genes codificadores de proteínas, como por exemplo, as glicoproteínas variantes de superfície (VSG) e proteínas PARP (*procyclic acidic repetitive protein*) de *T. brucei* (78, 79).

Nos eucariotos a transcrição ocorre da seguinte maneira: a RNA polimerase II liga-se ao gene promotor iniciando a transcrição do gene (80). O processo acaba resultando no transcrito primário ou pré-RNA, contendo sequências de íntrons e éxons. Com a adição de 7-metilguanossina (m7G cap) na extremidade 5' do RNA, ocorre a remoção dos íntrons por *cis-splicing* e adição da cauda de adenina (poli-A) na extremidade 3', gerando o mRNA maduro (73, 81). Por outro lado, a transcrição em tripanossomatídeos (**Figura 4**) ocorre pelo reconhecimento das regiões promotoras tanto pela RNA polimerase I quanto a RNA polimerase III (78, 82-84). Porém, para a maioria dos genes, não existe sequência consenso para a polimerase II. Sendo a RNA polimerase II de baixa especificidade e devido a ambas as fitas do DNA poderem ser transcritas (72), a RNA polimerase inicia a transcrição indiscriminadamente no genoma e forma longos RNAs policistrônicos sem íntrons (85). Com isso,

esses pré-RNAs policistrônicos são processados por poliadenilação na extremidade 3' e *trans-splicing* na extremidade 5' pela adição de uma sequência líder de aproximadamente 39 nucleotídeos chamada de *spliced leader* (**Figura 4**) (86). Ao contrário do que ocorre nos procariotos, nos tripanossomatídeos as unidades monocistrônicas maduras codificam proteínas que não pertencem a uma mesma via metabólica e os níveis de mRNAs são modulados por eventos pós-transcricionais (73).



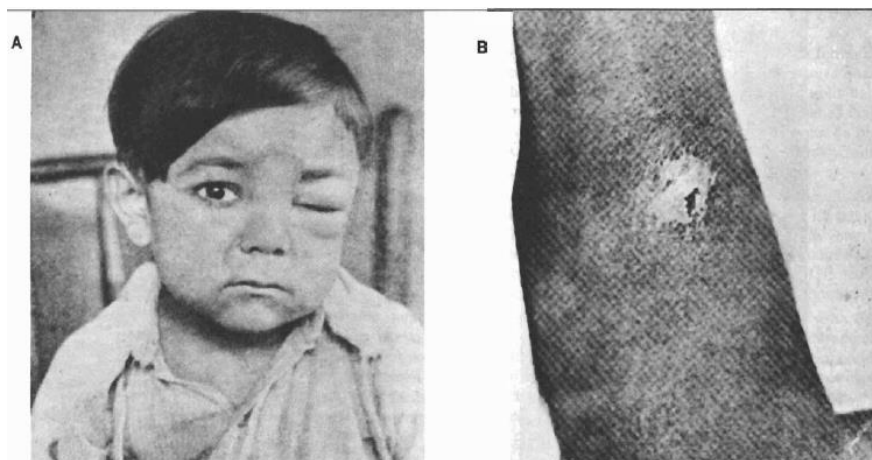
**Figura 4** – Representação da maquinaria de transcrição e processamento de mRNAs em tripanossomatídeos. Os genes do genoma são transcritos como pré-RNAs policistrônicos que por sua vez são processados por poliadenilação na extremidade 3' e *trans-splicing* na extremidade 5' pela adição de *spliced leader*, gerando mRNAs monocistrônicos maduros (87).

## 1.8 Doença de Chagas

### 1.8.1 Fase Aguda

Esta fase inicia-se após a infecção das formas metacíclicas em hospedeiros humanos. A infecção pode ocorrer pela mucosa dos olhos (sinal de Romaña) ou por ferimento causado pelo inseto ou pelo ato de coçar (chagoma de inoculação) (**Figura 5**). Caso a contaminação ocorra por via oral, congênita ou por transfusão sanguínea não haverá sinal aparente (88). A fase aguda é normalmente assintomática, porém manifestações não específicas podem ser observadas, como febre, náusea, vomito, anorexia, miocardite, linfadenopatia e hepatoesplenomegalia. Esses sintomas iniciam-se de 1 a 2 semanas após a picada do inseto infectado, ou após alguns meses se a transmissão ocorrer por transfusão de sangue infectado. Após esse período, os anticorpos IgM são encontrados em níveis elevados no soro dos pacientes portadores da DC (89) e também são

detectados anticorpos IgG nessa fase (90). O diagnóstico é realizado em menos de 10% dos casos, devido aos sintomas não específicos, e mais de 95% dos casos curam-se espontaneamente (91).



**Figura 5** – Via de entrada. a) Sinal de Ronaña; b) Chagoma de inoculação. Fonte: Neves DP et al (92).

### 1.8.2 Fase Crônica

Após a fase aguda, 60-70% dos pacientes entram na fase latente, na qual não há desenvolvimento dos sinais clínicos da doença. Essa fase pode durar anos ou a vida inteira do paciente. Os outros 40-30% desenvolvem a fase crônica, que caracteriza-se por comprometimento cardíaco ou digestivo (megacólon e megaesôfago) como revisto em Rassi A Jr (4). Além disso, A DC pode comprometer o sistema nervoso central e periférico (93). Essa fase pode iniciar-se 10-30 anos após o contato com o parasito.

### 1.8.3 Tratamento

O tratamento da doença de Chagas iniciou-se nas décadas de 60-70 com a utilização do Nifurtimox (Lampit®, Bayer) e do Benzonidazol (Rochagan® ou Rodoni®, Roche), ambos fármacos de escolha para o tratamento da fase aguda da doença (94). O modo de ação do Nifurtimox se dá pela redução metabólica de grupos nitro por nitronucleases, que por sua vez, leva a produção de radicais nitroânios. Como o mecanismo de desintoxicação do *T. cruzi* é ineficiente, uma diminuição no oxigênio leva a morte do parasito. A utilização do Nifurtimox foi descontinuada na década de 80 no Brasil, Chile e Argentina devido a seus efeitos colaterais. Já o modo de ação do Benzonidazol se dá por ligações covalentes ou outras interações intermediárias de nitroredução de componentes do parasito, ou por ligação ao DNA, lipídeos ou proteínas do parasito (95). O Benzonidazol também apresenta efeitos colaterais e não é recomendado para pacientes grávidas, com insuficiência cardíaca e renal, infecções sistêmicas, neoplasias, idosos e pessoas com baixa imunidade (95). Os

tratamentos usuais contra o *T. cruzi* ainda estão abaixo do ideal, necessitando o desenvolvimento de novas abordagens.

### 1.9 Diagnóstico da Doença de Chagas

O diagnóstico da doença de Chagas na fase aguda normalmente não ocorre, devido aos sintomas dessa fase serem inespecíficos, prejudicando o tratamento da doença (90). O teste para a fase aguda baseia-se em exames parasitológicos diretos, nos quais uma grande quantidade de tripomastigotas pode ser visualizada por microscopia utilizando sangue fresco com anticoagulantes (96). Para o diagnóstico da fase crônica, é realizado o teste sorológico para detectar anticorpos contra *T. cruzi*, por exemplo, testes de IFI (Teste de Imunofluorescência Indireta), HAI (Teste de Hemaglutinação Indireta) e o teste ELISA (Ensaio de Imunoabsorção Enzimática). Esses testes apresentam alta sensibilidade, porém baixa especificidade por apresentar reação cruzada com *Leishmania sp.* e *Trypanosoma rangeli* (97). A PCR (Reação em Cadeia da Polimerase) não é utilizada para diagnóstico clínico e sim em pesquisa laboratorial, devido à falta de padronização dos métodos (4). O diagnóstico da doença de Chagas congênita é realizado utilizando-se o cordão umbilical ou sangue venoso do recém-nascido, sendo um método muito invasivo (98). Embora os métodos apresentados sejam utilizados rotineiramente no laboratório para o diagnóstico da doença de Chagas, nenhum deles apresenta sensibilidade e precisão de 100%, tornando necessário o desenvolvimento de novas metodologias para a caracterização da doença.

### 1.10 Proteômica Baseada em Espectrometria de Massas

Em 1994 durante a “*First Siena Conference, 2D Electrophoresis: From Maps to Genomes*” surgiu o termo “PROTEOME”, cuja autoria foi dada a Marc R. Wilkins, Vitaliano Pallini e Denis Hochstrasser (99, 100). Além do mais, o termo “*The PROTEin complemente expressed by a genOME*” foi definido por Wilkins (100). Em outras palavras, o proteoma pode ser definido como um conjunto de proteínas expressas por um genoma em um determinado tempo, espaço, estado (patológico ou não) e estímulos externos (101, 102), definindo diferentes proteomas para a mesmo genoma.

O estudo do transcriptoma (conjuntos de transcritos) pode analisar o estado pontual da célula em determinado tempo, condição, modificações pós-transcricionais ou *splicing* alternativos, porém, essas análises não irão fornecer informações sobre as proteínas expressas pelo fato de ocorrerem inúmeras modificações pós-traducionais (fosforilação, glicosilação, ubiquitinação, arginilação, etc) (102, 103). A proteômica é utilizada como ferramenta para a caracterização

detalhada do proteomas, suas modificações pós-traducionais (PTMs), quantificação de todas as proteínas de uma amostra como também a interação entre proteínas (104).

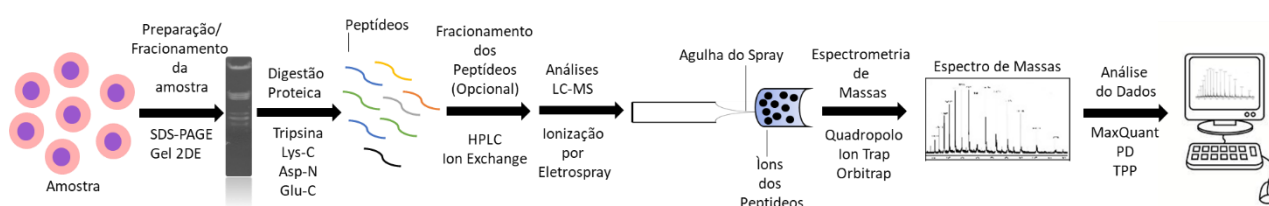
O surgimento da proteômica como a conhecemos hoje teve início na década de 1980, com a utilização da separação de proteínas por gel bidimensional (*two-dimensional gel electrophoresis, 2DE*) (105, 106). Essa técnica consiste na separação de proteínas utilizando dois parâmetros, o ponto isoelétrico (carga) e a massa molecular, e ambos são definidos baseados na mobilidade das proteínas em uma matriz de gel de poliacrilamida (106). Mesmo sendo o 2DE uma metodologia poderosa para identificação de proteínas, a mesma tem limitações, como por exemplo, a dificuldade em identificar proteínas de membranas, estando limitado a detecção de proteínas associadas a membranas e proteínas com baixa hidrofobicidade (107).

O estudo do proteoma de tecidos, células ou organismos inteiros são conduzidos na maioria dos laboratórios utilizando a estratégia de *bottom-up*, também denominada *shotgun*. Esse método baseia-se na caracterização de proteínas de uma dada mistura pela análise de seus peptídeos gerados através de proteólise enzimática (108, 109). O nome *shotgun* foi dado por Yates pela analogia do sequenciamento genômico (101). A abordagem *shotgun* permitiu a determinação de proteomas e tem sido largamente utilizada por sua alta reprodutibilidade, sobressaindo-se ao método clássico de análise 2DE (110). A vantagem sobre os géis 2DE se dá pela limitação desta última técnica na identificação de proteínas que compartilham determinadas características, como por exemplo, hidrofobicidade acentuada, ponto isoelétrico extremo e alta massa molecular (111, 112). Sendo assim, proteínas importantes podem deixar de ser identificadas caso as análises sejam baseadas apenas na separação em géis (113, 114). Na estratégia *bottom-up* (**Figura 6**) amostra de proteínas de interesse sofre proteólise por tripsina, passa por um processo de dessalinização e injetadas no espectrômetro de massas (explicado mais detalhadamente adiante).

A proteômica foi beneficiada com o desenvolvimento tecnológico dos equipamentos para análise qualitativa e quantitativa de proteínas e peptídeos. Anteriormente, o maior desafio era o desenvolvimento de métodos que permitissem que peptídeos pudessem ser ionizados e detectados sem sua destruição nesse processo (99). Essa limitação foi ultrapassada com o surgimento de novas metodologias para o sequenciamento de peptídeos utilizando metodologias denominadas *soft ionization*, como o ESI (Ionização por Electrospray), concebida por John Fenn, e MALDI (Ionização por Dessorção a Laser Assistida por Matriz) por Koichi Tanaka, Franz Hillenkamp e Michael Karas, acompanhado pela miniaturização e automação da cromatografia líquida (Cromatografia Líquida, LC) (115), permitindo quantificação e identificação de peptídeos com maior velocidade e precisão (116). Adicionalmente, com a rápida melhora nos equipamentos de cromatografia e espectrômetros de

massas e o desenvolvimento de algoritmos mais precisos nas análises dos dados gerados, a proteômica baseada em espectrometria de massas tornou-se uma ferramenta valiosa no estudo das proteínas.

Com a maior utilização de métodos de separação livre de géis, foram desenvolvidos novos métodos de separação baseados na cromatografia líquida acoplada a espectrometria de massas em tandem (LC-MS/MS). A proteômica baseada em espectrometria de massas tem ajudado nas análises de sistemas biológicos e a cada dia que passa vão surgindo novas melhorias nas metodologias empregadas nas análises em proteômica. De fato, a proteômica auxilia na compreensão de processos biológicos e nos últimos anos tem ajudado na descoberta de muitos potenciais biomarcadores, embora poucos deles de fato passaram para a fase de clínica (117-119). Alguns detalhes importantes da proteômica dizem respeito ao desenho experimental, na qual sofre grande influência a preparação da amostra, isto é, qual é o melhor experimento a se fazer para responder à pergunta em questão. Outro detalhe muito importante é a análises dos dados, necessitando de pessoas qualificadas e treinadas (116). A **Figura 6** mostra as a diversas estratégias que podem ser utilizadas na proteômica baseada em espectrometria de massas.



**Figura 6** – Visão geral da proteômica baseada em espectrometria de massas *bottom-up*. Adaptado de Steen H, Mann M (120).

### 1.10.1 Avanços Tecnológicos na Proteômica Baseada em Espectrometria de Massas

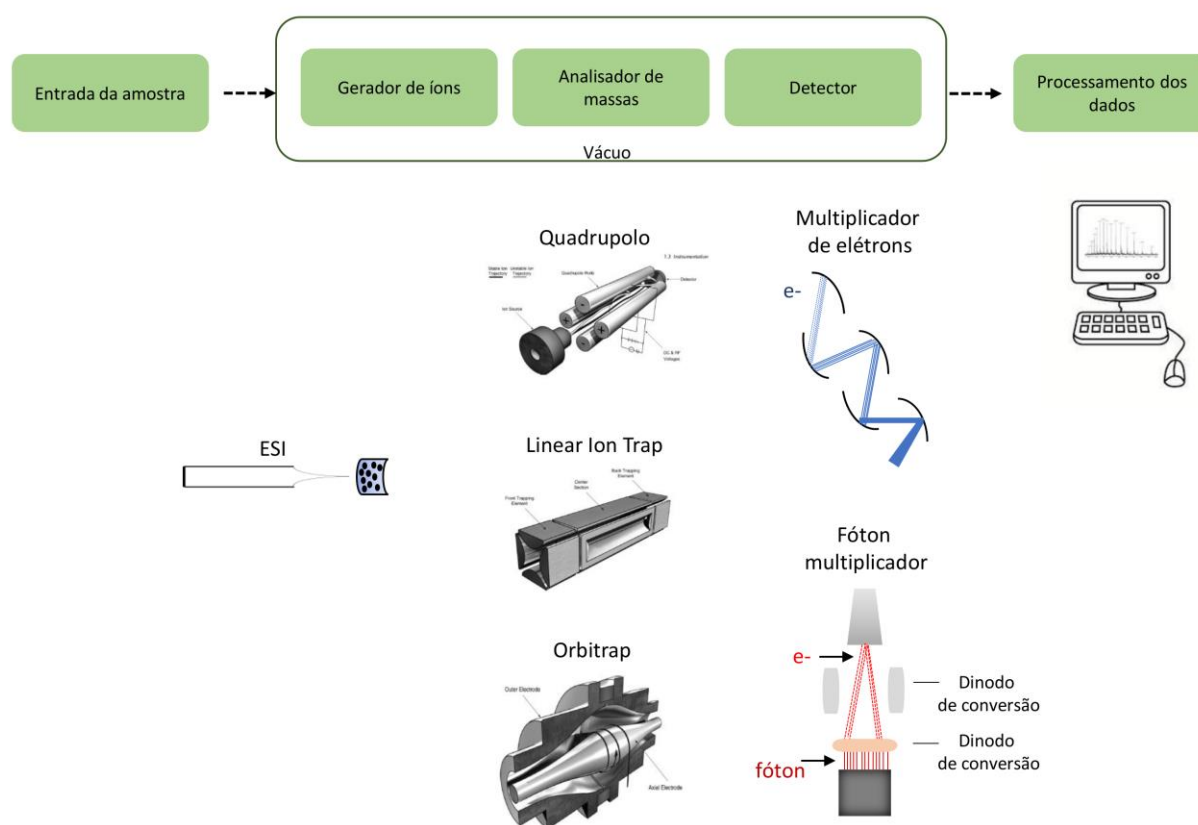
Um espectrômetro de massas tem quatro partes fundamentais: a fonte de ionização, o analisador de massas, o detector e o sistema de aquisição de dados (mostrado na **Figura 7**). Primeiramente a amostra a ser analisada é introduzida na fonte de ionização. No caso do ESI as moléculas (peptídeos) serão dissolvidas em solução polar e volátil, para então serem bombardeadas através de um fino capilar de metal (que é mantido entre 3 a 5 kV), e nebulizadas na ponta do capilar com aplicação de forte campo elétrico, permitindo a formação de pequenas gotículas carregadas. As gotículas carregadas são evaporadas por aplicação de calor e nitrogênio seco e as cargas residuais são transferidas para os peptídeos (evento denominado *droplet fission*) (121). Logo em seguida, os peptídeos ionizados são conduzidos através de campo elétrico até o analisador de massas. É possível



manipular a fragmentação aumentando a tensão, e conseqüentemente aumentando a colisão com as moléculas de nitrogênio (122).

No analisador de massas ocorre a separação de acordo com as suas respectivas razões massa/carga ( $m/z$ ) (123). Os analisadores de massas nessa fase da análise têm grande influência na identificação dos peptídeos, isto porque, cada analisador de massa tem parâmetros diferentes tais como, resolução, sensibilidade e exatidão diferentes uns dos outros, implicando na qualidade do dado adquirido (espectro) (124). Os analisadores de massas mais utilizados na pesquisa proteômica atualmente são: Tempo de voo (TOF); Quadrupolo; Armadilha de íons (*Ion Trap*, IT) e Orbitrap (123-129). O elemento final do espectrômetro é o detector, cuja função será a de registrar os dados de abundância relativa e os valores de  $m/z$  dos íons, e apresentá-los como espectros de massas (130).

A **Figura 7** sumariza todos os componentes do espectrômetro de massas.



**Figura 7** – Componentes do espectrômetro de massas. Fontes de íons ESI. Diferentes modelos de analisadores de massas e os diferentes métodos de detector. Módulo de processamento de dados. Adaptado de Hart-Smith G, Blanksby SJ (131).

### 1.10.2 Fragmentação de Peptídeos

Na proteômica utiliza a espectrometria de massas para analisar peptídeos e identificar proteínas em banco de dados. Um dos primeiros métodos desenvolvidos, que usa a informação da massa molecular dos peptídeos (digeridos enzimaticamente), foi o PMF (*Peptide Mass Fingerprint*). Com o surgimento dos analisadores híbridos (dois analisadores acoplados), permitiu-se o

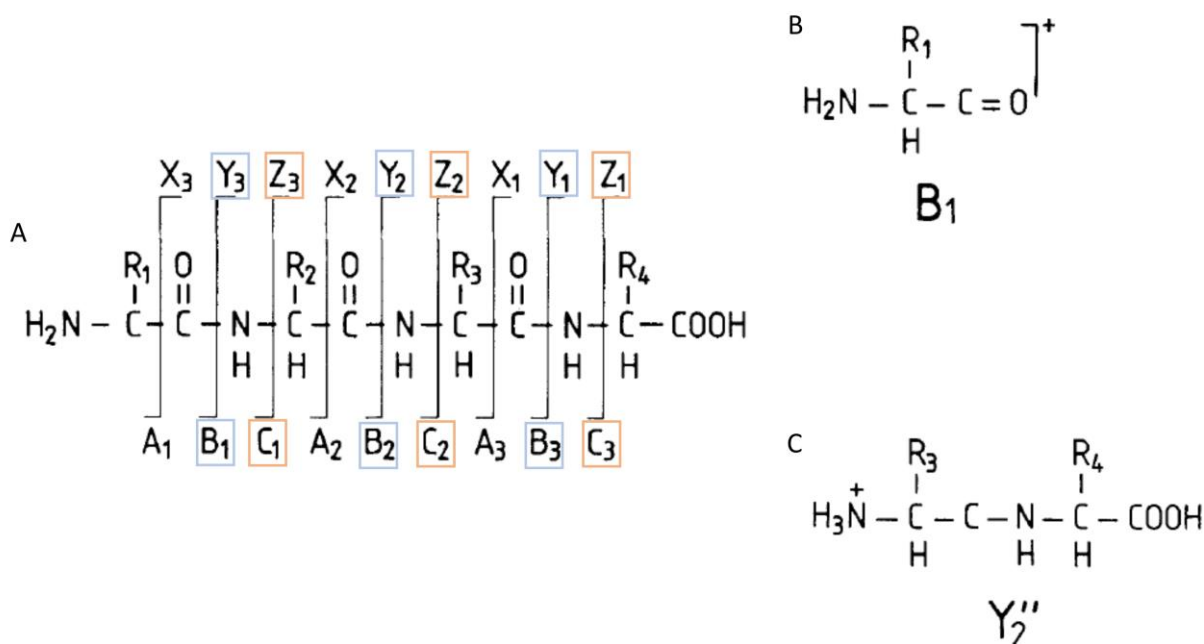
surgimento de experimentos em sequência (*tandem*), sendo então possível detectar um determinado íon e submetê-lo a uma etapa de fragmentação (101, 132).

Na etapa de fragmentação um método que é bastante utilizado é a dissociação induzida por colisão (CID)(133). O processo de fragmentação começa com os íons sendo introduzidos na região do vácuo do espectrômetro de massas (por meio de ESI ou MALDI). Em seguida, os íons precursores são acelerados seguidos de colisões em gás neutro (hélio, argônio ou nitrogênio), levando a conversão de energia cinética em energia vibracional. Conforme a energia vibracional excede um limiar, as ligações peptídicas covalentes podem quebrar. Os íons fragmentos resultantes são usados para obter a sequência peptídica em busca em banco de dados.

Seguindo o modelo de mobilidade do próton (134), que descreve como a energia adquirida induz a transferência intramolecular dos prótons em cada peptídeo, culminando na desestabilização das ligações do esqueleto polipeptídico, e por consequência, induzindo a formação de dos íons fragmentos (135), os íons fragmentos são classificados como íons que retêm a carga residual (próton) no lado N-terminal (gerando fragmentos  $-a$ ,  $-b$  e  $-c$ , dependendo da ligação fragmentada) e íons que retêm a carga residual (próton) na região C-terminal (gerando os fragmentos  $-x$ ,  $-y$  e  $-z$ , dependendo da ligação que é fragmentada), segundo a nomenclatura proposta por Roepstorf-Fohlman-Biemann (136) (**Figura 8**). Outro detalhe é a formação do íon denominado imônio: esse íon formado da clivagem simultânea das regiões C- e N-terminais de um resíduo de aminoácido. Esse íon é útil para diagnóstico fornecendo informações do íon precursor (99).

Outros métodos de fragmentação utilizados são a dissociação por captura de elétrons (*Electron Capture Dissociation*, ECD) e a dissociação de transferência de elétrons (*Electron Transfer Dissociation*, ETD) (137, 138).

Recentemente um método de fragmentação baseado na dissociação induzida por alta energia de colisão (HCD) vem sendo muito utilizado (139). Esse método foi disponibilizado comercialmente pela empresa Thermo® (LTQ-Orbitrap™) em 2007. Com um processo de fragmentação semelhante ao CID, os íons são primeiramente fragmentados na câmara de colisão, em seguida são enviados ao C-trap (que é um LIT curvado), que comprime eletrodinamicamente os íons no tempo e espaço e os acelera, através de um gradiente rápido de voltagem para serem ejetados para o orbitrap (140). A fragmentação do HCD gera íons  $b$  e  $y$ , podendo gerar outros íons menores (**Figura 8**). O HCD permite observar íons de baixo  $m/z$ , tornando-o adequado para análises iTRAQ (*Isobaric Tags for Relative and Absolute Quantitation*) ou TMT (*Tandem Mass Tags*) (141-144).



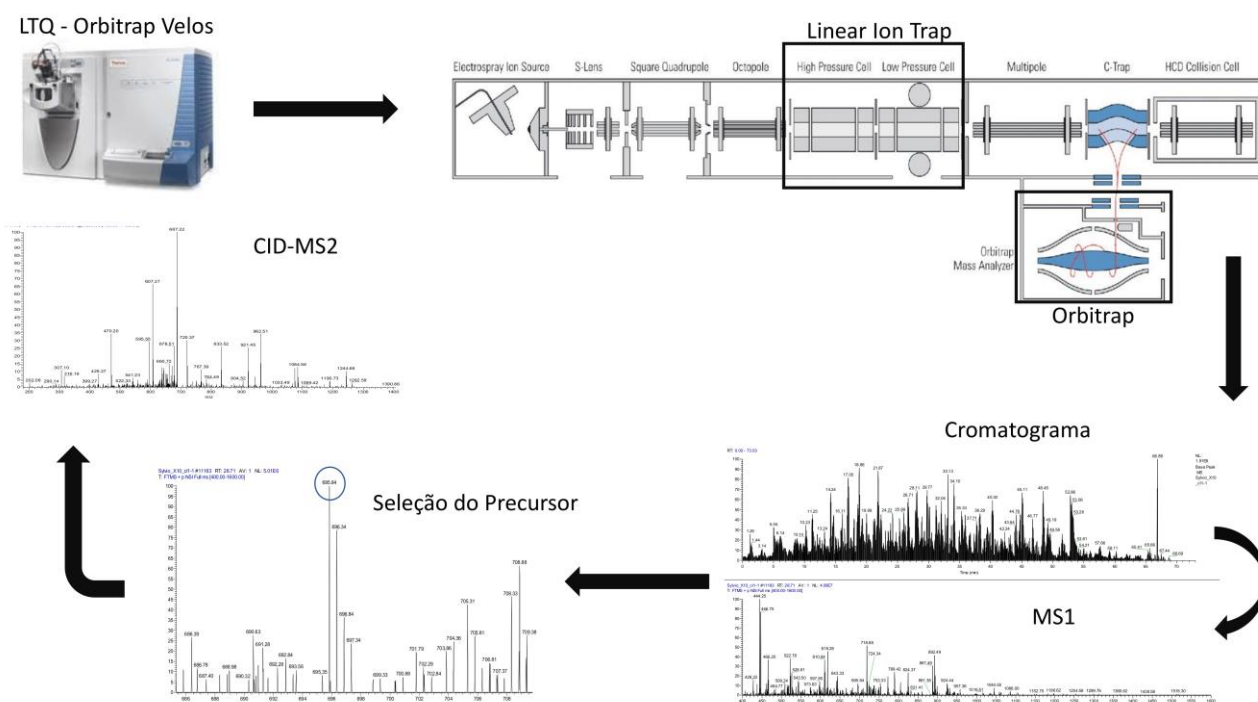
**Figura 8** – Representação dos íons gerados nos diferentes métodos de fragmentação. Quando a fragmentação ocorre clivagem no N-terminal os fragmentos são chamados de A, B e C, e quando ocorre no C-terminal do peptídeo os fragmentos são chamados de X, Y e Z. (A) – A numeração indica qual ligação peptídica é clivada contando a partir do N e C, respectivamente. (B) – As cargas a direita significam que houve um rearranjo de cargas no processo de fragmentação, esses íons são denominados B. (C) – Os íons Y são os fragmentos mais comuns gerados pela clivagem da região C-terminal. Laranja: Fragmentação CID e HCD; Azul: Fragmentação ECD e ETD. Adaptado de Roepstorff P, Fohlman J. (136).

A espectrometria de massas sequencial (MS/MS) consiste em vários processos: isolamento das amostras de interesse, digestão enzimática, ionização por um gás neutro, seleção de massa do íon precursor, dissociação (por exemplo CID), detecção dos íons e aquisição dos dados. Em especial, ela surge como uma alternativa para a obtenção de maior seletividade nos experimentos quantitativos.

Quando o experimento em espectrometria de massas MS/MS ocorre em regiões separadas do aparelho (como no triplo quadrupolo), diz-se que é espectrometria de massas sequencial no espaço (*tandem mass spectrometry in space*). Porém, quando todos esses processos ocorrem no mesmo espaço físico (como no *ion trap*), nomeia-se como espectrometria de massas sequencial no tempo (*tandem mass spectrometry in time*), permitindo a realização de vários estágios de  $\text{MS}^n$  (145).

Dependendo do objetivo do trabalho, espectrometria sequencial oferece quatro tipos de análises: varredura de íons produtos, varredura de íons precursores, varredura de perda neutra e monitoramento seletivo de reações. Nesse trabalho foi utilizado um espectrômetro de massas LTQ-Orbitrap Velos com aquisição dependente de dados (DDA) (**Figura 9**). Esse tipo de análise é baseado na análise de varredura de íons precursores (*product ion scanning*): o primeiro analisador seleciona

o íon de interesse ( $MS^1$ ) que serão fragmentados na câmara de colisão, e os íons fragmentos serão analisados no segundo analisador de massas ( $MS^2$ ) (146, 147).



**Figura 9** - Visão geral do funcionamento do LTO-Orbitrap Velos em modalidade de aquisição dependente de dados (DDA). O funcionamento do LTO-Orbitrap Velos baseia-se na análise de varredura de íons precursores, o primeiro analisador seleciona os íons de interesse ( $MS^1$ ), que serão fragmentados na câmara de colisão, e os íons fragmentos serão analisados no segundo analisador de massas ( $MS^2$ ). Para esse exemplo foi utilizada a cepa *Tc Sylvio X10 CL1*.

### 1.10.3 Análise de Dados Proteômicos

Após todo o procedimento de preparação da amostra, processamento no espectrômetro de massas e obtenção dos espectros  $MS/MS$ , a próxima etapa é realizar a busca em banco de dados de proteínas. Inicialmente, é necessário realizar um pré-processamento e para isso existem métodos que classificam espectros  $MS/MS$  em bons ou ruins, levando em consideração a boa relação sinal-ruído, fragmentação uniforme e extração precisa das massas de cada pico no espectro de massas. Para isso, utilizam-se algoritmos que são responsáveis pela classificação dos espectros antes de submete-los à busca em banco de dados (148).

Um dos softwares para buscas em bancos de dados utilizando espectros  $MS$  são o Sequest, Mascot e o Andromeda.

O Sequest, que usa um modelo descritivo, é baseado na previsão de como os peptídeos se fragmentam no modo em tandem ( $MS/MS$ ), que é então quantificado para determinar a qualidade da combinação entre a predição e o espectro experimental (149).

O Mascot e Andromeda usam o modelo probabilísticos. Resumidamente, o modelo probabilístico determina a probabilidade da relação entre os espectros MS/MS e as sequências no banco de dados (149).

O MaxQuant utiliza o algoritmo Andromeda, que está disponível gratuitamente ([www.maxquant.org](http://www.maxquant.org)), sendo essa uma das vantagens em relação ao Mascot. O Maxquant pode ser executado em computadores mais modestos. Outra vantagem é sua capacidade de analisar peptídeos co-fragmentados, possibilitando a análise de espectros MS/MS de alta resolução, melhorando as identificações principalmente em misturas complexas (150).

### 1.11 Biblioteca Espectral

A busca de sequência proteicas em bancos de dados usando espectros MS/MS é um processo computacionalmente caro, propenso a erros e muito trabalhoso. Ela tem sido objeto de pesquisa desde os primórdios da proteômica (151, 152). Muitas ferramentas foram desenvolvidas no intuito de alcançar o equilíbrio entre sensibilidade e precisão (152), mas mesmo com a evolução do hardware e software, as buscas ainda continuam dispendiosas necessitando de grandes processamentos quando se quer analisar grande conjuntos de dados, deixando essas análises para grandes grupos de pesquisas. (152).

O conceito de uso da biblioteca espectral inicia-se por volta da década de 70, quando a estrutura de uma espécie química era inferida pelo seu padrão de fragmentação gerado pelo espectrômetro de massas (153). Esse método envolvia curagem manual e era demorada.

O uso inicial de bibliotecas espectrais foi em análises de pequenas moléculas, em particular na cromatografia gasosa acoplada a plataforma de espectrometria de massas (GC-MS) empregando ionização eletrônica (EI) que fragmenta os analitos na fonte de íons. As bibliotecas construídas eram distribuídas pelos fornecedores de espectrômetros de massas e também para outras entidades como o Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (*National Institute of Standards and Technology*, NIST- USA) (154). Essa biblioteca era formada por uma grande coleção de espectros de massas teóricos de compostos conhecidos. Logo em seguida, um espectro de massa gerado experimentalmente podia ser comparado com a biblioteca buscando a melhor correspondência possível. O único impedimento da sua utilização na época era a disponibilidade de bibliotecas confiáveis (154, 155).

No âmbito da proteômica, uma idéia semelhante de busca espectral foi apliado por Yates et al em 1998 (155, 156). Yates et al obtiveram sucesso no uso de uma “correlação cruzada” normalizada para medir a semelhança entre um espectro CID e um espectro de referência obtido do

mesmo instrumento. Mesmo que essa ideia tenha sido revelada bem a frente do seu tempo, bibliotecas espectrais de peptídeos não existiam antes de 1998, lançando bases para as buscas em bibliotecas espectrais como conhecemos hoje (157).

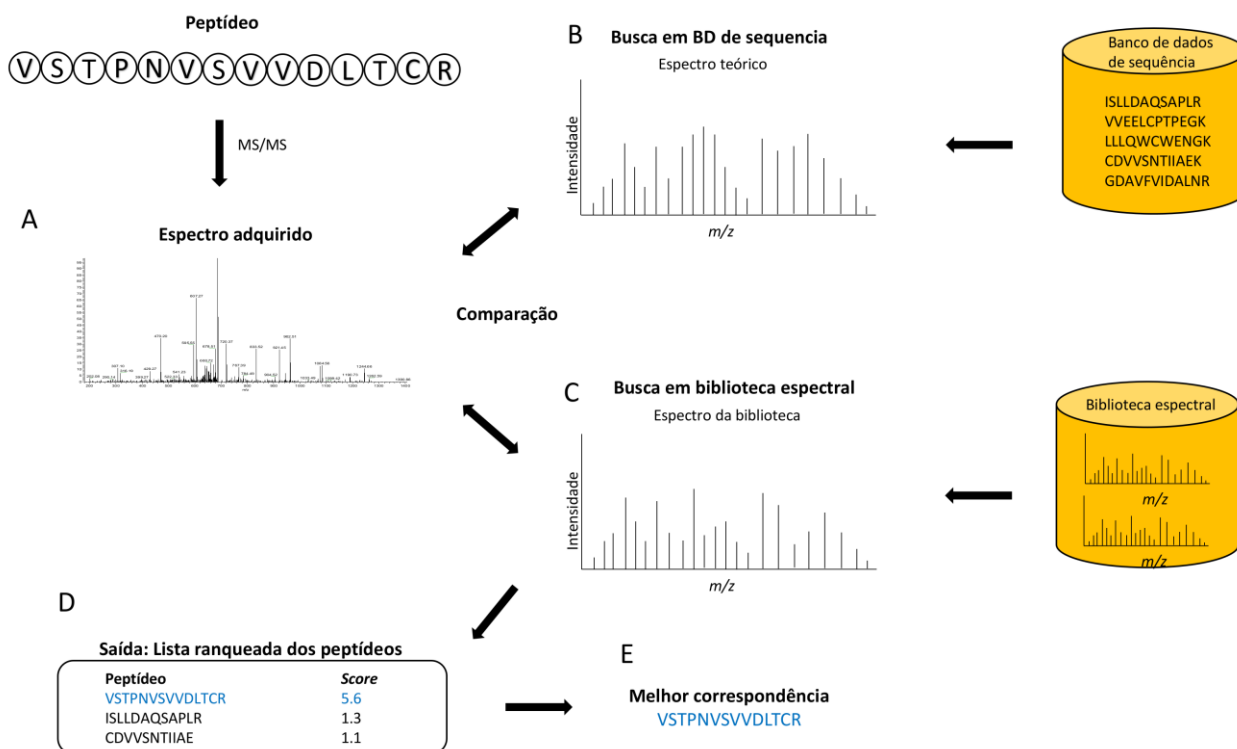
Com a evolução dos equipamentos e dos métodos de preparações de amostras, atualmente a proteômica pode rever métodos utilizados anteriormente (156, 158). Além disso, tem aumentado a disponibilidade on-line dos dados gerados por vários laboratórios, e os dados proteômicos *shotgun* tem sido padronizados (159-163). Todavia, para que a biblioteca espectral seja abrangente é importante construí-la com espectros MS/MS de uma grande variedade de fontes para prover precisão, qualidade e evitar os falsos positivos. Para tal, o NIST- USA em colaboração com o projeto PeptideAtlas estão responsáveis pela qualidade dos espectros MS/MS disponíveis on-line (160, 164).

A busca espectral é uma forma de agregar conhecimento de experimentos anteriores em novos estudos, já que normalmente esses dados gerados ficariam sem uso. Assim, a busca espectral é ideal para a proteômica direcionada, isto é, quando não se buscam novas informações e sim informações repetitivas de um certo conjunto pré-definido de peptídeos (152, 165). Os softwares mais usados para busca espectral são o SpectraST (*Spectra Search Tool*) (152, 166), X!Hunter (158) e Bibliospec (156), (**Figura 10 e 11**).

O Bibliospec desenvolvido por Frewen et al em 2006 (156), e mostraram que é possível comparar espectros adquiridos de diferentes modelos de equipamentos em diferentes laboratórios. No mesmo ano foi apresentado o X!Hunter (161) por Craig et al, que utiliza abordagem semelhante para inferir a significância estatística de correspondências espectrais. Por volta de 2007 Lam et al revela o Spectrast (152, 166) que foi integrada na suíte da Trans-Proteomic Pipeline (167-169), uma suíte completa para análise de dados proteômicos que possibilita a realização de todas as etapas de análise de dados desde a conversão de formato de dados até a validação estatística.

A busca por biblioteca espectral vem adquirindo importância nesses anos e para explicá-la melhor, é importante compará-la com o método de busca em banco de dados de sequência. A busca em banco de dados de sequência de proteínas é frequentemente o método de escolha para identificar espectros de massas. Quando se compara um espectro de massas MS/MS com espectros previstos teoricamente de peptídeos com precursor  $m/z$  semelhantes, o algoritmo de busca de banco de dados de sequências atribui a sequência a melhor correspondência ao espectro de massas observado como identificação putativa (157). Desde que os peptídeos estejam incluídos no banco de dados de sequências e que as modificações sejam consideradas nas buscas, o algoritmo é teoricamente capaz de identificar os peptídeos. Por essa razão, a busca de banco de dados de sequências é ideal para descobrir novos peptídeos que não tenham sido observados

experimentalmente. Por outro lado, a busca em bibliotecas espectrais é menos utilizada para esse propósito, porque não depende do genoma para a identificação, e esse aspecto fica restrito a identificação de peptídeos previamente identificados (157).



**Figura 10** - Identificações dos peptídeos podem ser realizadas comparando os espectros experimentais contra B - um banco de dados de espectros teóricos contidos em um banco de dados de sequências proteicas ou C – Busca em biblioteca espectral. D – Lista dos melhores candidatos obtidos e E – Melhor correspondência dos dados experimentais contra um banco de dados de sequência proteica. Adaptado de Nesvizhskii AI (170).

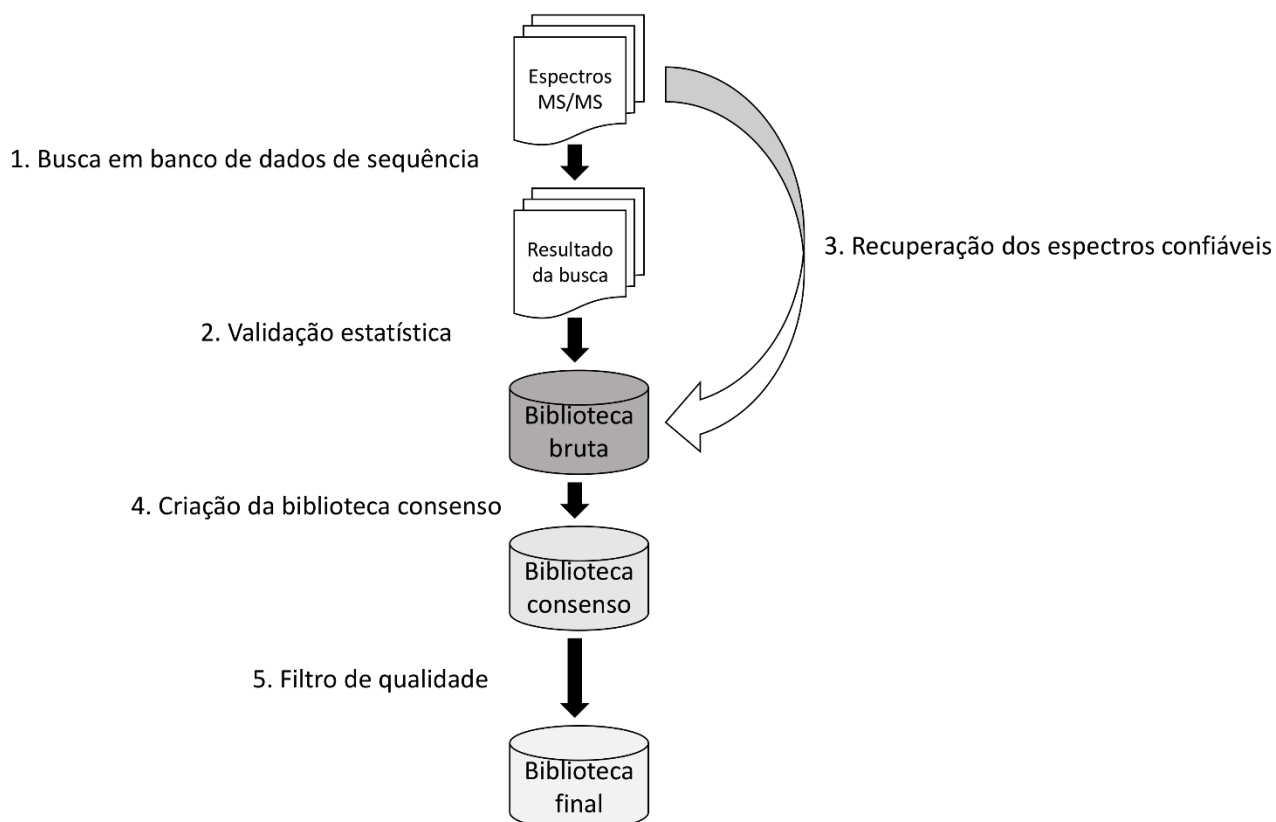
Porém, as buscas em bibliotecas espectrais apresentam algumas vantagens em relação a busca em banco de dados de sequências. A primeira delas é o princípio de correspondência espectral, e não requer qualquer informação prévia sobre os fragmentos do analito ou do espectrômetro de massas. Além do mais, a busca em biblioteca espectral faz uso de toda as informações contidas no espectro de massas para distinguir entre boas e más comparações, significando que a sensibilidade é geralmente maior do que a busca em banco de dados. Segundo, a biblioteca espectral representa uma fração do proteoma do organismo estudado, com isso, o “espaço de busca” é muito menor em comparação com busca em banco de dados. Este fato revela não só uma melhora na velocidade de busca como também a sensibilidade, reduzindo os falsos negativos (155, 158). Terceiro, a biblioteca espectral conecta experimentos passados com futuros, enquanto que a busca em banco de dados de sequências não traz conhecimentos prévios no processo de identificação (157). Se o objetivo do experimento não é descobrir peptídeos novos, mas sim identificar o mesmo peptídeo em diferentes

amostras, a busca por biblioteca espectral é mais adequada do que busca por banco de dados de sequências. E por fim, as limitações mencionadas na busca por biblioteca espectral podem ser parcialmente superadas pelo aumento da biblioteca espectral e também pela utilização de métodos de previsão de espectros.

A construção de uma biblioteca espectral MS/MS pode ser resumidamente dividida em cinco passos. O primeiro passo envolve a análise dos espectros por ferramentas tradicionais de buscas de bancos de dados de proteínas. No segundo passo, é realizada a validação estatística para dar confiabilidade às identificações. A terceira etapa envolve a consolidação de todas as identificações, incluindo a tarefa de integrar dados e resultados de diferentes formatos e lugares, que é realizada manualmente. No quarto passo, quando a biblioteca “bruta” é construída, os espectros com várias identificações são unidos, formando-se um espectro de consenso, isto é, um “consenso” para esse peptídeo. Esse passo diminui a redundância da biblioteca. Neste passo pode-se usar uma abordagem alternativa: alternativamente a formar um espectro de consenso, seleciona-se o “melhor” espectro que representa as repetições. E por fim, no quinto passo é realizado o controle de qualidade, ou seja, nesse passo os espectros erroneamente identificados ou ruidosos são removidos da biblioteca (171).

De modo geral, a busca em biblioteca espectral inclui três etapas para a identificação correta. A primeira etapa é o pré-processamento, nesta etapa o ruído e outros picos que oferecem baixa discriminação serão removidos do espectro de consulta. O propósito do pré-processamento é reter a informação discriminativa utilizando por exemplo, picos intensos e discriminativos que são úteis para a correspondência espectral. Para tal, as intensidades de picos são transformadas por alguma função matemática (por exemplo o produto escalar), para reduzir a importância relativa de combinar intensidades de picos, que estão sujeitas a variações experimentais. A segunda etapa é a quantificação do grau de similaridade entre o espectro de consulta e os espectros da biblioteca, essa quantificação pode ser por comparação do espectro de consulta com cada um dos espectros candidatos da biblioteca. A etapa final é a validação estatística, ou seja, atribuir uma probabilidade à identificação, que é considerada a combinação espectral mais próxima (157).





**Figura 11** – Fluxograma para a construção de uma biblioteca espectral. 1 – Após todo o procedimento de obtenção de espectros MS/MS realiza-se uma busca em banco de dados de sequência. 2 – Após a busca, valida-se os peptídeos com testes estatísticos. 3 – Integra-se todos os dados de diferentes lugares e formatos. 4 – Espectros semelhantes são unidos em um consenso. 5 – Aplicação de filtros de qualidade para prover melhor qualidade da biblioteca espectral. Adaptado de: Lam, H. (172).

O uso de biblioteca espectral está expandindo consideravelmente. Mais recentemente, a metodologia já foi utilizada para identificar fontes de alimentos de carrapatos (173), microrganismos (174) (como bactérias) e neste caso protozoários. A busca espectral é um recurso muito útil para pesquisas em proteômica principalmente quando o objetivo é identificar os mesmos peptídeos em diferentes amostras (171).

### 1.12 Proteômica no *Trypanosoma cruzi*

Devido a capacidade da proteômica baseada em espectrometria de massas em identificar e quantificar proteínas e também suas PTMs (104), é interessante aplica-la para o estudo do *T. cruzi* já que seus diferentes estágios de desenvolvimento provavelmente têm expressão diferencial de proteínas.

Seguindo essa linha de raciocínio, o primeiro estudo realizado em *T. cruzi* utilizando LC-MS/MS foi feito em 2004 por Paba et. al. (175). Neste estudo foram identificadas 1573 proteínas de 1710 espectros MS/MS. Uma análise mais rigorosa resultou na identificação de 41 proteínas com no

mínimo 2 peptídeos por proteína. Das proteínas identificadas, 7 estavam relacionadas a organização celular, 7 ao metabolismo e o restante pertenciam a destinos proteicos. Além disso, 29 proteínas apresentaram expressão conservada e 9 proteínas foram mais expressas em tripomastigotas e 3 em amastigotas (175). A maioria das proteínas encontradas pertenciam a classe de proteínas de choque térmico (*Heat-Shock Proteins*, HSPs). Para validar os resultados obtidos deste estudo e de um estudo anterior, na qual foi utilizada a metodologia gel 2DE (176), Paba et al., utilizando metodologia ICAT (*Isotope-Code Affinity Tag Technology*, ICAT), encontraram taxas similares entre as duas metodologias.

Posteriormente, em um trabalho realizado por Atwood et al. envolvendo a análise do proteoma do *T. cruzi* durante seu desenvolvimento (177) foram gerados 139.147 espectros MS/MS. Desse total de espectros foram identificados 1168 grupos de proteínas contendo 2784 proteínas totais e 5.720 peptídeos únicos de alta confiabilidade. Das 2784 proteínas identificadas, 1008 são anotadas como proteínas hipotéticas. Adicionalmente, nas análises das diferentes formas do parasito, foram identificadas 1871 proteínas totais na forma amastigota, 1486 em tripomastigotas, 2339 em tripomastigotas metacíclicos e 1861 em epimastigotas. A melhora nas identificações de proteínas no trabalho de Atwood et al., em comparação com o trabalho de Paba et al., provavelmente se deu devido a publicação do genoma do *T. cruzi* em 2005 (64), melhorando consideravelmente as identificações.

Com o objetivo de caracterizar a expressão de proteínas na metaciclologênese, Parodi-Talice et al. (178) utilizaram géis 2DE e MALDI-TOF. Com a utilização dos géis 2DE foram obtidos 500 spots de géis, das quais foram selecionados 100 spots para análises em MALDI-TOF. Dos 100 spots de géis, 20 spots apresentaram baixa qualidade dos espectros de massas. Os 80 spots restantes apresentaram ótima qualidade, sendo que 66 spots (82%) foram identificados em buscas em bancos de dados correspondendo a 43 proteínas diferentes. Além disso, as proteínas foram agrupadas em categorias funcionais, sendo 33% relacionadas com atividades metabólicas, 16% relacionadas com transporte de elétrons e atividades antioxidantes, 16% com biossíntese, proteólise e catabolismo, 12% estavam relacionados com proteínas estruturais, 14% foram identificadas como proteínas hipotéticas, os restantes 9% reuniam proteínas como as calmodulinas, adenilato ciclase e proteínas de superfície como a glicoproteína GP90. Em particular, os autores sugerem que as PTMs durante a metaciclologênese podem ser de grande importância na regulação da expressão gênica durante a diferenciação e que a glutamato desidrogenase (GluDH) pode desempenhar um papel importante durante o estresse oxidativo por aumentar a captação de aminoácidos como fonte de nutrientes.

Um dos primeiros estudos com o objetivo de relacionar filogeneticamente as seis DTUs utilizando a proteômica foi realizado pelo grupo de Telleria et al. (179). Para tal, foi utilizada a metodologia 2DE-DIGE com MALDI TOF-TOF em um conjunto de 26 clones de *T. cruzi* representando as seis DTUs. Os resultados do trabalho sugerem que algumas proteínas podem ser especificamente associadas a diferentes DTUs, ou sua expressão diferencial podem ser correlacionadas. A maioria das proteínas identificadas são estruturais, como as  $\alpha$ - e  $\beta$ -tubulinas.

Visando a aplicação clínica, Diaz et al. analisaram e identificaram a expressão diferencial das formas tripomastigotas e amastigotas de isolados clínicos da DTU-I (180). Para esse estudo os autores utilizaram géis 2DE para a obtenção do perfil proteico e MALDI-TOF ou LC-MS/MS para a identificação das proteínas. As análises revelaram que em média, os géis 2DE revelaram 325 spots em cada forma do parasito. Comparando as duas formas do parasito, foram encontrados 33 spots de peptídeos exclusivos de tripomastigota e 28 de amastigota. Além disso, 21 spots de géis foram expressos com mais intensidade em tripomastigotas, enquanto que 30 spots foram mais intensos em amastigotas. As 16 proteínas estavam associadas principalmente ao metabolismo glicolítico e enquadravam-se em proteínas estruturais.

Outros trabalhos foram realizados para obter diferentes frações do *T. cruzi*, como por exemplo, o trabalho de Queiroz et. al. (181) que avaliaram o subproteoma das formas epimastigotas do *T. cruzi* utilizando duas metodologias. A primeira baseando-se na tripsinização da superfície celular (*Shave*) das células vivas intactas, e a segunda abordagem utilizava a biotilação das proteínas de superfície seguido de isolamento cromatográfico de afinidade com estreptavidina. Ambos os métodos foram analisados por LC-MS/MS. Os resultados obtidos mostraram que as duas metodologias são complementares e abrangentes, revelando muitas informações sobre o subproteoma do parasito. Adicionalmente, em outro trabalho desenvolvido pelo grupo de Queiroz et. al. (182), utilizando-se da mesma metodologia, porém estudando outras formas do parasito (tripomastigota e amastigota axênico) os autores mostraram um grande repertório de proteínas no subproteoma da membrana plasmática, revelando possíveis alvos para fármacos. Além disso, a análise de bioinformática mostrou que a maioria das proteínas preditas estão envolvidas na infecção das células hospedeiras, na adesão celular e na modulação da resposta imune do mamífero.

Em um estudo para analisar a fração organelar da forma epimastigotas do *T. cruzi* conduzido por Ferrel et. al. (183) baseando-se na metodologia de Scott e Docampo (184) com utilização da LC-MS/MS, os autores identificaram 396 proteínas, das quais 138 foram anotadas como hipotéticas. Uma análise comparativa com um estudo de proteoma de células inteiras resultou na validação de 173 novas proteínas, sendo 38 proteínas não encontradas em estudos anteriores em epimastigotas.

A conclusão a que os autores chegaram é que o enriquecimento de frações subcelulares pode ajudar na detecção de novas proteínas que não são encontradas em estudo proteômicos em larga escala. Além da fração organelar, um outro estudo foi realizado com a fração nuclear de células de epimastigotas de *T. cruzi*, em que Santos et. al. (185) adaptaram duas metodologias (186, 187) para isolar a fração nuclear do parasito. As análises por 2D-nLC-MS/MS (*Two-Dimensional nano-Scale Liquid Chromatography*) revelaram 12875 peptídeos de 864 proteínas. Esse foi o primeiro estudo proteômico em larga escala realizado de fração nuclear. Dentre as 864 proteínas, 272 foram anotadas como não caracterizadas, e 275 não foram encontradas em análise global do proteoma do *T. cruzi*. Para confirmar o método desenvolvido, análises de bioinformática foram realizadas, revelando 65 cluster de genes, em que foi mostrado o agrupamento de membros com organização de cromatina e funções ligadas ao DNA.

Todos esses estudos reforçam a importância da proteômica como ferramenta de apoio para outras metodologias. Sabendo que as manifestações da DC podem estar relacionadas com a variabilidade genética dos parasitos (58, 188), o desenvolvimento de uma nova ferramenta que utiliza espectros de massas para caracterização de DTUs em *T. cruzi* associada a utilização de ferramentas de bioinformática, ajudará a determinar qual DTU está envolvida em uma determinada região de um modo mais rápido comparado com os testes rotineiros.

Além disso, a espectrometria de massas pode ser uma importante ferramenta para caracterização das diferentes cepas de *T. cruzi*, pelo fato de ser uma tecnologia robusta, sensível, precisa e simples.

Os métodos recentes para a caracterização das DTUs contêm limitações, como por exemplo, erro na interpretação das análises (61). Outros métodos não são simples para serem utilizados para uma caracterização rápida, principalmente quando há grandes coleções de isolados (59). Por outro lado, podemos utilizar a espectrometria de massas para complementar análises moleculares ajudando a caracterizar as DTUs e, conseqüentemente levando a um melhor diagnóstico e abordagem da doença no paciente. Tendo em vista que cada DTU tem um grau de virulência (58), a caracterização das diferentes DTUs pode levar a um melhor diagnóstico da doença.

Este é o primeiro estudo de classificação das DTUs de *T. cruzi* utilizando somente a informação dos espectros MS/MS, sem nenhum conhecimento prévio sobre a sequência proteica.

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Desenvolver um método de caracterização das diferentes cepas de *T. cruzi* (DTUs) utilizando espectrometria de massas e ferramentas de bioinformática.

### 2.2 Objetivos Específicos

1. Desenvolver uma plataforma baseada na análise de espectrometria de massas para discriminar as diferentes cepas de *T. cruzi*;
2. Desenvolver métodos computacionais para gerar banco de dados de espectros de massas de fragmentação das diferentes DTUs;
3. Validar o método proposto através de espectros de fragmentação (MS/MS) obtidos das amostras não conhecidas com o banco de dados;
4. Identificar e quantificar as sequências de peptídeos/proteínas expressas em cada cepa por meio de *database searching* em banco de dados;
5. Analisar quais redes de interação proteína-proteína estão reguladas nas diferentes DTUs.

### 3 MATERIAL E MÉTODOS

#### 3.1 Cultura Celular de *Trypanosoma cruzi*

As formas epimastigotas de *T. cruzi* foram cultivadas em meio LIT (*Liver Infusion Tryptose*) suplementado com 10% soro fetal bovino em pH 7.2 a 28 °C (189), em fase exponencial. Neste estudo foram utilizadas apenas cepas sequenciadas e validada (**Tabela 2**).

**Tabela 2** – Lista de cepas utilizadas para o desenvolvimento e validação do método Tc-STAMS<sup>2</sup>.

DTUs	Tripanossomatídeo	Hospedeiro	Localidade
I	<i>T. cruzi sylvio X10/4</i>	<i>Homo sapiens</i>	Belém, PA, BR
I	<i>Sylvio X10 CL1</i>	<i>Homo sapiens</i>	Pará, Brasil
I	<i>T. cruzi G mucurae</i>	<i>Didelphis marsupialis</i>	Manaus, AM, BR
II	<i>T. cruzi Y</i>	<i>Homo sapiens</i>	Rio Grande do Sul, RS, BR
II	<i>T. cruzi esmeraldo</i>	<i>Homo sapiens</i>	São Felipe, Bahia, BR
III	<i>T. cruzi M-6241 Cl 6</i>	<i>Homo sapiens</i>	Belém, PA, BR
III	3869	<i>Homo sapiens</i>	Amazonas, Brasil
III	<i>T. cruzi</i>	<i>Homo sapiens</i>	Carauari, AM, BR
IV	<i>T. cruzi Can III Cl 1</i>	<i>Homo sapiens</i>	Belém, PA, BR
IV	<i>T. cruzi Jose Julio</i>	<i>Homo sapiens</i>	Amazonia, AM, BR
V	<i>T. cruzi NR Cl 3</i>	<i>Homo sapiens</i>	Bolívia
V	<i>T. cruzi 92:80 Cl 2</i>	<i>Homo sapiens</i>	Bolívia, Santa Cruz
V	MN cl2	<i>Homo sapiens</i>	Region IV, Chile
VI	<i>T. cruzi CL 14</i>	<i>Triatoma infestans</i>	Rio Grande do Sul, BR
VI	<i>T. cruzi CL Brener CL 1</i>	<i>Triatoma infestans</i>	Rio Grande do Sul, BR
Tc bat	<i>T. cruzi bat Cl 1.1</i>	<i>Myotis levis</i>	São Paulo, SP, BR
Outros	<i>T. marinkellei</i>	<i>Carollia perspicillata</i>	Monte Negro, RO, BR
	<i>T. dionisii</i>	<i>Epitesicus brasiliensis</i>	São Paulo, SP, BR
	<i>T. erneyi</i>	<i>Mops condylurus</i>	Moçambique, Chupanga, Sofala
	<i>T. rangeli AM-80</i>	<i>Homo sapiens</i>	Amazônia, AM, BR

Nota - Tabela com todas as informações a respeito das DTUs utilizadas neste estudo.

### 3.2 Influência das Condições de Cultivo do *Trypanosoma cruzi*

As células da cepa *Sylvio X10/1* foram coletadas em fases exponencial e estacionária de crescimento. Duas réplicas biológicas para a fase estacionária (St1 e St2) e duas para a fase exponencial (Exp1 e Exp2) foram analisadas.

### 3.3 Preparação das Amostras e Análise nLC-MS/MS

As formas epimastigotas ( $5 \times 10^8$  células em fase exponencial) foram lavadas três vezes em PBS (*Phosphate-Buffered Saline*) pH 7.2 (8,000 g por 10 minutos em temperatura ambiente), foram resuspendida em 400  $\mu$ L de tampão de lise (7 M Uréia, 2 M Tiouréia, 1 mM DTT (*DL-Dithiothreitol* – SIGMA-Aldrich) e inibidores de protease (Amersham) e foram incubadas em agitação por 30 minutos para a solubilização das proteínas. As proteínas foram reduzidas com DTT 10 mM final, alquiladas com iodo-acetamida 40 mM final (Sigma-Aldrich), digeridas com tripsina (Promega) na proporção 1:50 ( $\mu$ g tripsina/ $\mu$ g proteínas) em 50 mM final de solução bicarbonato de amônio por 16 horas a 37 °C. A reação foi interrompida com ácido fórmico 1% (pH menor que 3) e em seguida, as amostras foram dessalinizadas em colunas de C18 (*StageTips*). O *StageTips* ou extração *stop-and-go* é um método de dessalinização simples que é bastante utilizado em laboratórios de proteômica (190). Foram preparadas quatro réplicas para cada DTU. Para a validação de diferentes laboratórios, as amostras A (DTU-III) e B (DTU-I) foram preparadas usando um mínimo de três réplicas de acordo com o protocolo descrito acima.

#### 3.3.1 Preparação da Amostra para as Condições Ácidas e Básicas

Os peptídeos foram dessalinizados em condições ácidas (ácido 1 e ácido 2) e básicas (básico 1 e básico 2). Em particular, para a purificação ácida, os peptídeos trípticos foram acidificados com 0,1% de TFA (pH 3) e carregados em uma microcoluna *StageTip* ativada com ácido antes de serem eluídos com 50% de acetonitrila com 0,1% de ácido trifluoroacético (TFA). Para a purificação básica, os peptídeos trípticos foram dissolvidos em bicarbonato de amônio a 0,1% (pH 10) e carregados em uma *Stage-tips* de base ativada antes de serem eluídos com acetonitrila a 50% com bicarbonato de amônio a 0,1%. Os peptídeos eluídos foram secos e analisados por espectrometria de massa.

### 3.3.2 Análises Nano LC-MS/MS

Os peptídeos foram separados em colunas Reprosil-Pur C18-AQ (3 $\mu$ m; Dr. Maisch GmbH, Germany) usando Easy-LC nano\_HPLC (Proxeon, Odense, Dinamarca). O gradiente utilizado foi 0-34% de solvente B (A = 0.1% de ácido fórmico; B = 90% ACN, 0.1% de ácido fórmico), em 70 min com um fluxo de 250 nL/min. As análises MS/MS foram realizadas utilizando o LTQ-Orbitrap Velos (Thermo Scientific, Bremen, Alemanha). As aquisições dos espectros de massas foram feitas em 400-1500  $m/z$  com resolução de 30,000. Em cada espectro de massa os 20 íons mais intensos foram selecionados para fragmentação utilizando CID no *ion trap* linear. Os parâmetros para aquisição dos dados foram: tempo de ativação = 15 ms, energia normalizada = 35, Q-ativação = 0,25, exclusão dinâmida = 1 repetição, tempo da exclusão = 30 segundos e limiar de intensidade = 30,000, íons alvo =  $2e^4$  como descrito por Palmisano et. al, (191).

### 3.3.4 Quantidades de Amostra, Gradientes Cromatográficos e Tipos de Fragmentação MS/MS

*Utilizados no Teste de Conceito.*

A robustez do método Tc-STAMS<sup>2</sup> foi testada usando diferentes parâmetros: 1) quantidade de amostra, 2) gradientes cromatográficos e 3) técnicas de fragmentação. Diferentes quantidades (0,5 e 1  $\mu$ g) e concentrações de amostras (*Low* e *High*) foram carregadas na coluna analítica antes da análise MS. O tempo de eluição cromatográfica foi ajustado para 20, 70 e 130 minutos de 0-34% de solvente B em 250 nL/min. A fragmentação CID foi utilizada para desenvolver o método Tc-STAMS<sup>2</sup>. O HCD foi avaliado como tipo de fragmentação peptídica alternativa em peptídeos trípticos separados por gradiente cromatográfico de 70 minutos. Para a fragmentação HCD, cada varredura de MS foi adquirida na resolução 30.000 FWHM (*Full Width at Half Maximum, FWHM*) seguido de 7 scans MS/MS de íons mais intensos com um tempo de ativação de 0,1 ms e energia de colisão normalizada de 35. A biblioteca espectral para cada DTU foi desenvolvida em um LTQ-Orbitrap Velos localizado no *PR Group* do Departamento de Bioquímica e Biologia Molecular da Universidade do Sul da Dinamarca (Odense, DK). Todos os outros testes para validar o método foram realizados em um LTQ-Orbitrap Velos (Thermo Fisher Scientific™) na unidade de espectrometria de massas Biomass-CEFAP (Centro de Facilidades para a Pesquisa) (São Paulo, Brasil). Para as análises adicionais foi utilizado o espectrômetro de massas Orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific™) disponível no Instituto de Química da Universidade de São Paulo. As configurações do aparelho foram: tempo de aquisição de 74 minutos a cada ciclo de MS/MS de 3 segundos. O MS<sup>1</sup> foi analisado no Orbitrap com resolução de 120,000 com alcance de  $m/z$  de 375-1600 com máximo tempo de injeção de 50 milissegundos. Os peptídeos foram analisados em modo positivo. Os espectros MS<sup>2</sup>



foram isolados no quadrupolo com janela de isolamento de 1,2 Da. Os peptídeos foram fragmentados via HCD com energia de colisão de 30, os íons fragmentos foram detectados no Orbitrap com resolução de 30,000 e tempo máximo de injeção de 54 milissegundos com AGC target de 50,000. No MS a primeira massa foi de 110. A intensidade mínima para os íons precursores serem selecionados foram de 50,000. Os íons precursores fragmentados foram excluídos por 45 segundos.

### 3.4 Bioinformática e Análises Estatísticas

#### 3.4.1 Geração da Biblioteca Espectral MS/MS e Correspondência Espectral

A geração da biblioteca espectral MS/MS e a comparação espectral foram realizadas usando o software SpectraST (versão 4.8) (152, 174, 175). Em particular, os espectros adquiridos por LC-MS/MS foram convertidos para um formato aberto (mzXML) pelo Msconvert (192), que faz parte da suíte oferecida pela TPP (*Trans-Proteomic Pipeline*) (167-169). O SpectraST foi utilizado para construir a biblioteca espectral e executar a correspondência espectral MS/MS (152). A biblioteca espectral de referência foi gerada com três arquivos “raw” (arquivos gerados pelo espectrômetro de massas) para cada DTU, e um arquivo *raw* de cada DTU foi utilizado para comparação com a biblioteca de referência para validação do método. A biblioteca espectral foi construída com as cepas das DTUs na forma epimastigota em fase exponencial. O primeiro passo para a geração da biblioteca espectral de referência envolve a aplicação de um limiar, dos quais os espectros MS/MS originados a partir do mesmo íon precursor peptídico são combinados para criar o espectro de consenso. Além disso, os espectros de baixa qualidade são excluídos da biblioteca (152). Para determinar a similaridade entre o espectro de consulta e a biblioteca, o SpectraST utiliza a função SDSS (*Spectral Dataset Similarity Score*) (168, 173). Em particular, o produto do ponto único (*unique dot product*) SDSS, foi abreviado como “score” ao longo do texto, e foi escolhida como função de discriminação entre as cepas de *T. cruzi* das DTUs (173). As confidências estatísticas nas identificações das DTUs corretas foram realizadas por bootstrap de dados (173, 193). Todas as ocorrências obtiveram um bootstrap de 1, a menos que fossem relatados bootstrap menores.

O software DiagnoProt (194) também foi utilizado para construção da biblioteca espectral e para a análise de similaridade. O DiagnoProt utiliza o mesmo método de comparação espectral, ele foi utilizado para validação computacional entre os diferentes softwares. Foram utilizados os parâmetros padrões para a criação do banco de dados espectral: Limite de similaridade de 0,70; tolerância ao precursor: 4,50; tipo de ativação: CID; número mínimo de picos: 50; deslocamento do compartimento 0,40; tamanho da caixa: 1,0005; mínimo bin  $m/z$  200,00; máximo bin  $m/z$ : 1700,00. O banco de dados espectral foi utilizado para combinar a identidade de cada amostra desconhecida.

### 3.4.2 Busca em Banco de Dados

Os arquivos *raw* MS/MS foram analisados usando: Proteome Discoverer, MaxQuant e Trans-Proteomic Pipeline. Com o Proteome Discoverer v2.1 (Thermo Fisher Scientific®) os espectros foram comparados com a base de dados de *T. cruzi* usando Mascot e Sequest. As buscas em banco de dados foram realizadas com os seguintes parâmetros: Tolerância de massa de precursor de 20 ppm; tolerância de massa MS/MS 0.5 Da (dados CID). A tripsina foi selecionada como enzima e carbamidometil cisteína como modificação fixa. As modificações das variáveis foram a oxidação da metionina e desamidação da glutamina e asparagina. As sequências peptídicas compartilhadas foram agrupadas como proteínas de acessos agrupadas (*grouped accessions proteins*). As taxas de falsos positivos (*False Discovery Rates*, FDR), foram realizadas usando o algoritmo Percolator com  $q$  igual ou inferior a 0.01. O FDR da proteína foi calculado no Proteome Discoverer e mantida abaixo de 1%.

Os arquivos *raw* também foram processados usando o MaxQuant (154) (versão 1.2.27.429). Os espectros MS/MS foram pesquisados utilizando o algoritmo Andromeda (150) contra o banco de dados de proteínas Uniprot do *T. cruzi* (11 de julho, 2017; 51.738 entradas). A tolerância de massa máxima permitida foi ajustada para 20 ppm para o precursor e depois ajustada para 4.5 ppm na busca principal, e para íons fragmentados foi de 0.5 Da. A especificidade enzimática foi ajustada para tripsina com um máximo de duas clivagens perdidas. A carbamidometilação cisteína (15.99 Da) foi fixada como modificação fixa, desaminação (NQ) e acetilação de proteínas N-terminal (42.01 Da) foram selecionadas como modificação variável. A análise bioinformática foi realizada utilizando o software Perseus v.1.5.2.6 (154), disponível no ambiente MaxQuant, e as sequências reversas e contaminantes foram excluídas nas análises posteriores. O FDR das proteínas foi calculado no MaxQuant e mantido abaixo de 1%. Os valores de intensidade do LFQ foram considerados para comparar relativamente a abundância de proteínas presentes nas diferentes DTUs.

A suite Trans-Proteomic Pipeline (TPP) foi utilizada para pesquisar arquivos *raw* convertidos para mzXML (167-169). Os arquivos mzXML foram pesquisados pelo algoritmo de busca Comet incorporado na plataforma do TPP (195). O FDR das proteínas e peptídeos foram estimados usando o algoritmo PeptideProphet incorporado ao ProteinProphet na TPP (196). A identificação com menos de 1% de FDR foi mantida.

Os dados *raw* de tecido humano (197) e *T. vivax* (Meta1, BSF1 e EP1) (198) foram obtidas do repositório público PRIDE (*PRoteomics IDentifications*), e utilizados como controle negativo.

### 3.4.3 Análise de Redes Interações Proteína-Proteína

Para as análises de interações proteína-proteína foi utilizado o software String (199) (<https://string-db.org/>). A tabela utilizada para essa análise está na **Tabela suplementar S6**. Para essas análises foram utilizadas a cepas *CL14* com *CL Brener* (ambas DTU-VI).

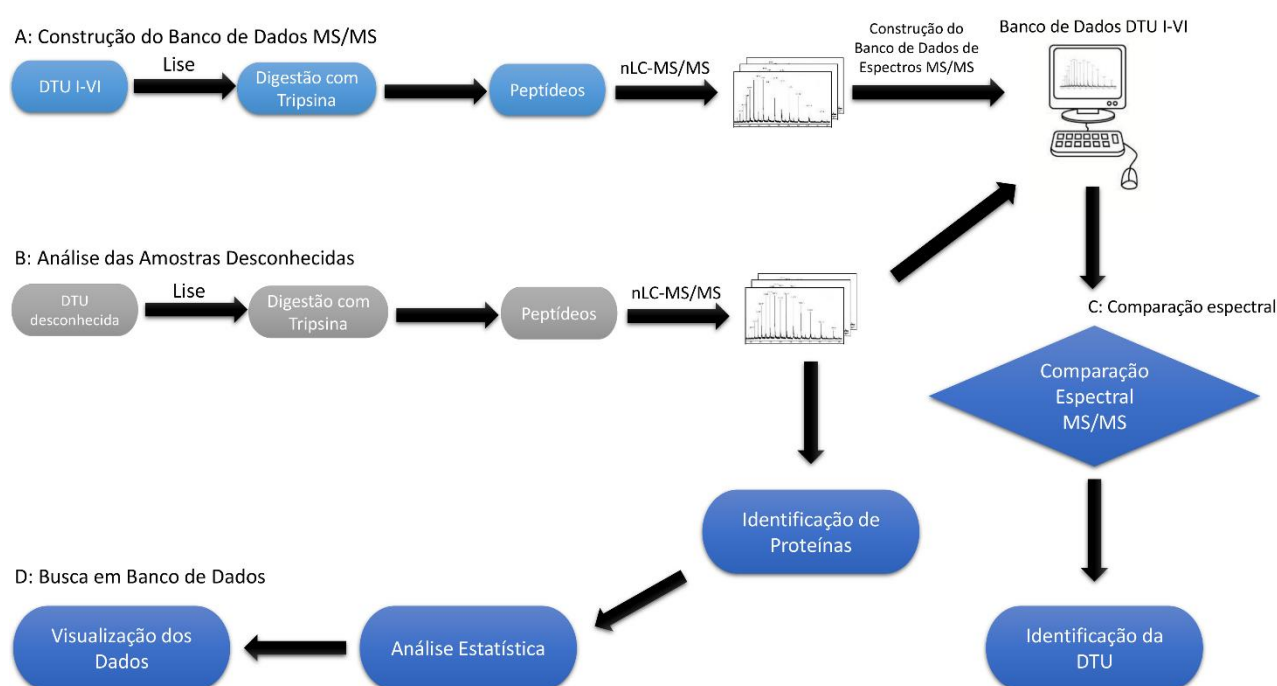
### 3.4.4 Análises Estatísticas

As quantificações livres de marcadores (*Label-Free Quantification*, LFQ) das proteínas e peptídeos foram quantificadas pelo Perseus (200). Valores significativamente regulados com valor de  $p$  menor que 0.05 e corrigidos com o teste Bonferroni post-hoc foram utilizados para agrupar diferentes DTUs. Os agrupamentos hierárquicos de proteínas e peptídeos significativamente regulados foram realizados usando o cálculo *Z-score* nos valores de intensidades de  $\log_2$  e foram representados como *heat maps*. As análises de componentes principais (PCA), foram realizadas utilizando o mesmo procedimento descrito acima no Perseus.

## 4 RESULTADOS

### 4.1 A Estratégia Tc-STAMS<sup>2</sup> Permitiu a Discriminação das DTUs

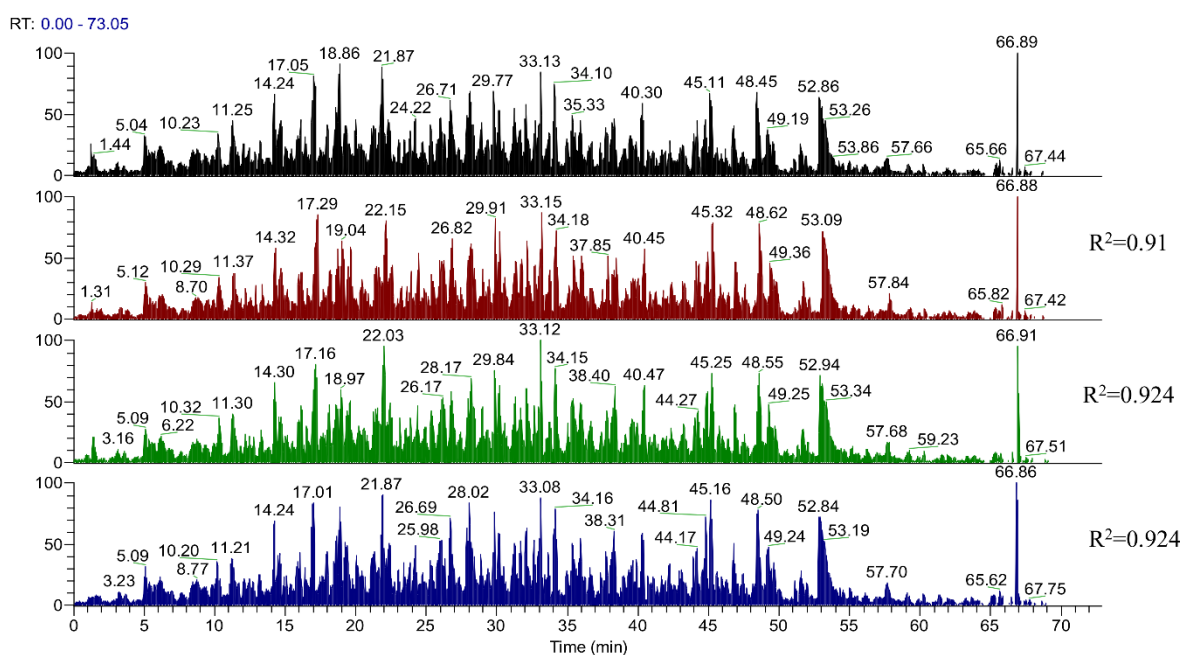
Neste estudo, a combinação de espectrometria de massas e abordagens computacionais foi utilizada para desenvolver um método para a discriminação de cepas de *T. cruzi* nas DTUs, chamado de ensaio de caracterização de cepas de *Trypanosoma cruzi* usando bibliotecas espectrais MS<sup>2</sup> de peptídeos (*Trypanosoma cruzi* Strain Typing Assay using MS<sup>2</sup>, Tc-STAMS<sup>2</sup>). Uma visão geral esquemática da identificação de *T. cruzi* nas DTUs utilizando espectros MS/MS de peptídeos trípticos é resumida na (Figura 12).



**Figura 12** – Tc-STAMS<sup>2</sup> fluxograma. (A) As proteínas foram extraídas da dorma epimastigotas, digeridas e submetidas a análise nLC-MS/MS. Os espectros MS/MS foram agrupados e mesclados para gerar uma biblioteca espectral de massa de referência utilizando o SpectraST (<http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST>) ou DiagonProt (<http://patternlabforproteomics.org/diagnoprot/>). (B) Os mesmos passos descritos no painel A foram utilizados para as amostras desconhecidas. Em particular, as DTUs foram lisadas, as proteínas foram digeridas utilizando tripsina. Os peptídeos foram analisados por nLC-MS/MS. (C) Os espectros gerados foram comparados com a biblioteca de referência para atribuir todas as amostras desconhecidas a uma DTU particular de *T. cruzi*. OS passos A, B e C foram utilizados em todas as análises de correspondência espectral. (D) Os dados obtidos pela espectrometria de massa MS/MS foram pesquisados contra a base de dados da proteína *T. cruzi*. Foram utilizados três plataformas computacionais: MaxQuant (154) Proteome Discoverer (Thermo Fisher) e Trans-Proteomic Pipeline (167-169).

Uma biblioteca espectral foi construída utilizando um total de 586513 espectros únicos MS/MS de peptídeos trípticos pertencentes a três arquivos *raw* de cada uma das seis cepas de *T. cruzi* (Figura 12 - A). Cada DTU foi processada e adquirida em quatro réplicas técnicas. Os espectros MS/MS adquiridos a partir de três réplicas de cada DTU foram utilizados para construir a biblioteca

espectral de referência usando o SpectraST (152). Após a construção da biblioteca MS/MS de referência, as proteínas da DTU desconhecida foram digeridas por tripsina antes de serem analisadas por nLC-MS/MS (**Figura 12-B**). Um perfil de LC-MS das quatro réplicas é mostrado na **Figura 13** e o *score* de correlação de Pearson indica alta semelhança entre as diferentes corridas. O perfil cromatográfico LC-MS dos peptídeos trípticos pertencentes as DTUs mostram alta semelhança entre as diferentes DTUs (**Figura 14**). Os espectros MS/MS das diferentes DTUs foram comparados contra a biblioteca usando o SpectraST (166). As identificações foram realizadas quando houve correspondências entre a biblioteca e o conjunto de dados espectrais das amostras testadas, neste caso, as diferentes DTUs. O resultado do SDSS foi utilizado para fornecer uma medida de similaridade quantitativa entre dois conjuntos de dados espectrais (abreviado de *score*) (**Figura 12-C**) (173). Foram realizadas buscas dos espectros MS/MS em banco de dados do proteoma do *T. cruzi*, conforme descrito na (**Figura 12-D**).



**Figura 13** – Perfil cromatográfico LC-MS dos peptídeos pertencentes as quatro réplicas da cepa *Sylvio X10 c11* (DTU-I). O índice de correlação de Pearson é relatado no lado direito do cromatograma. Utilizando a primeira réplica como referência, temos: A primeira réplica comparada com ela mesma resulta em  $R^2 = 1$ , segunda em  $R^2 = 0.91$ , terceira em  $R^2 = 0.924$  e a quarta em  $R^2 = 0.924$ .

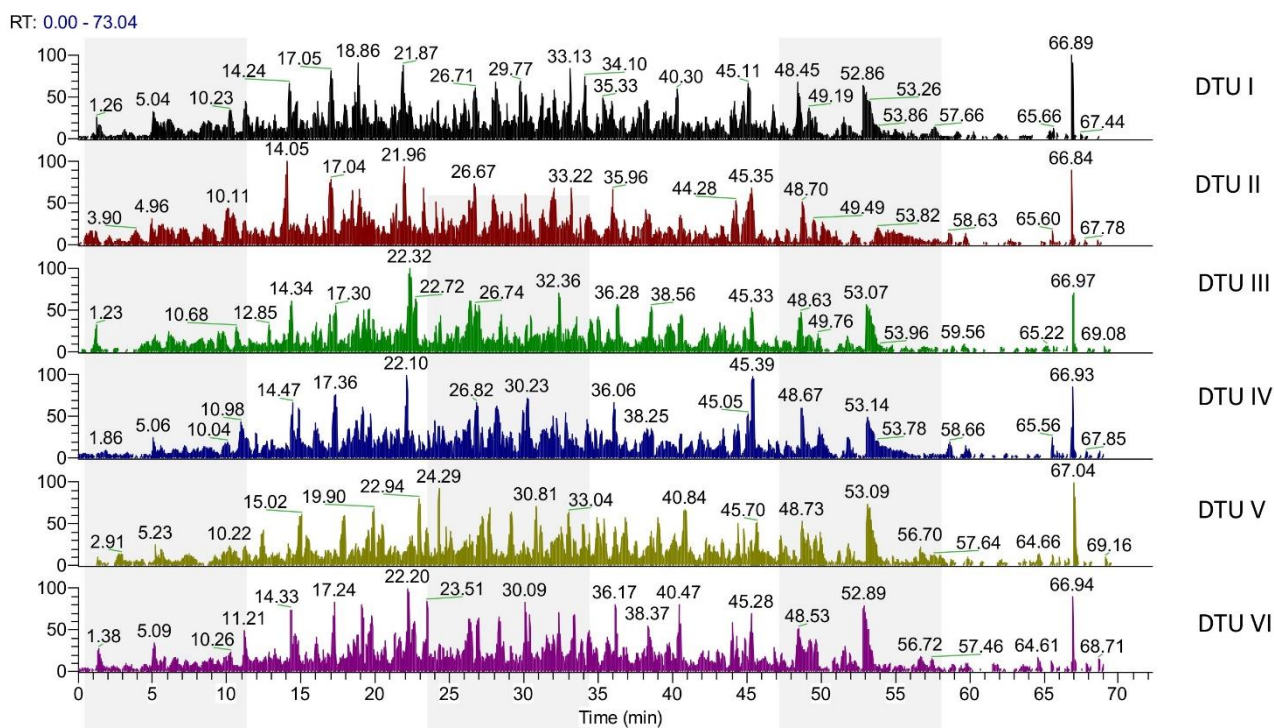


Figura 14 – Perfil cromatográfico LC-MS dos peptídeos pertencentes das seis DTUs *T. cruzi*.

Baseado no *score* de similaridade entre os espectros de massas MS/MS das amostras desconhecidas e a biblioteca espectral, o Tc-STAMS<sup>2</sup> foi capaz de diferenciar e identificar com precisão as diferentes DTUs. Nós observamos que o *score* ficou entre 0.75-0.86 para identificações corretas e 0 para casos de não identificações, como mostrado na Tabela 3.

**Tabela 3** - Identificações das DTUs baseadas nas buscas de similaridade espectral. As seis DTUs (*Sylvio X10 cl1*, *Y*, *M6241 cl6*, *CanIII cl1*, *MN cl2* e *CL Brener*). Os *scores* são relatados juntamente com o número de correspondências de espectro MS/MS. As similaridades são identificadas com alta pontuação e destacadas em cinza.

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
I	0.861 (2858/20232)	0.005 (11/13450)	0.008 (24/13419)	0.008 (26/14135)	0.016 (65/13470)	0.013 (33/14469)
II	0.010 (23/14212)	0.754 (1169/20004)	0.032 (72/14820)	0.023 (62/15126)	0.007 (22/14198)	0.020 (35/16038)
III	0.006 (18/13177)	0.020 (45/13838)	0.817 (2086/19587)	0.013 (42/13135)	0.017 (42/13689)	0.015 (35/14200)
IV	0.011 (38/14307)	0.013 (28/14213)	0.010 (32/13412)	0.849 (2648/20199)	0.004 (16/12889)	0.019 (48/14722)
V	0.008 (35/12560)	0.009 (20/12405)	0.013 (37/12771)	0.004 (16/11687)	0.863 (3086/19777)	0.011 (28/14215)
VI	0.016 (51/14230)	0.017 (27/14880)	0.013 (32/14101)	0.017 (51/14047)	0.022 (57/14849)	0.800 (1463/20238)

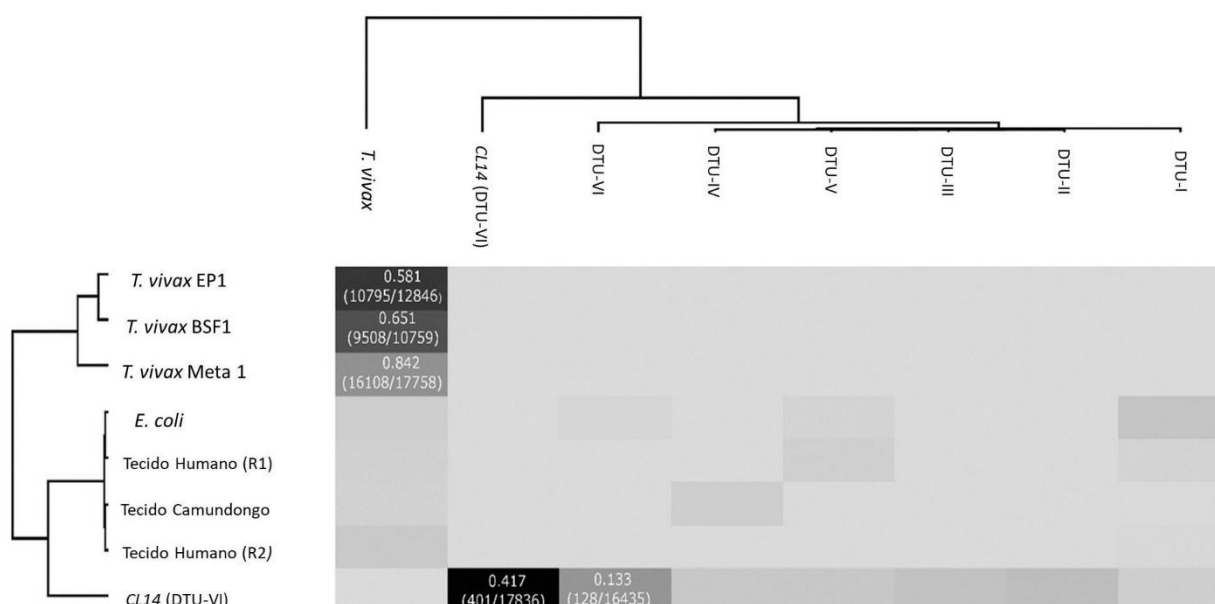
Os cromatogramas LC-MS obtidos para cada réplica e para cada DTU (**Figuras 13 e 14**) apresentam alta similaridade. No entanto, o método desenvolvido foi capaz de diferenciar e identificar cada um deles. Para excluir a possibilidade de que diferentes fases de crescimento pudessem influenciar a atribuição do algoritmo, foram coletadas formas epimastigotas *Sylvio X10/1* (DTU-I) na fase exponencial e estacionária. Independentemente da fase de crescimento, o algoritmo foi capaz de atribuí-lo à DTU correta (**Tabela 4**). Isso demonstra que o método de identificação não é afetado pelas fases do parasito.

**Tabela 4** - Comparação das correspondências espectrais MS/MS usando diferentes fases de crescimento da forma epimastigota de *T. cruzi Sylvio X10/1* (DTU-I).

Amostras	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI	CL14 (DTU-VI)
A (St1)	0.283	0.032	0.068	0.017	0.075	0.032	0.034
A (St2)	0.288	0.035	0.067	0.015	0.068	0.029	0.030
B (Ex1)	0.296	0.036	0.057	0.026	0.104	0.058	0.046
B (Ex2)	0.301	0.041	0.056	0.027	0.099	0.053	0.051

Nota - Tabela construída com *score* do SpectraST das identificações da forma epimastigota de *T. cruzi Sylvio X10/1* (DTU-I). Foi construída outra biblioteca com as seis DTUs e a cepa *CL14* da DTU-VI. St1 – Fase estacionária; St2 – Réplica biológica da fase estacionária; Ex1 – Fase exponencial; Ex2 – Réplica biológica da fase exponencial.

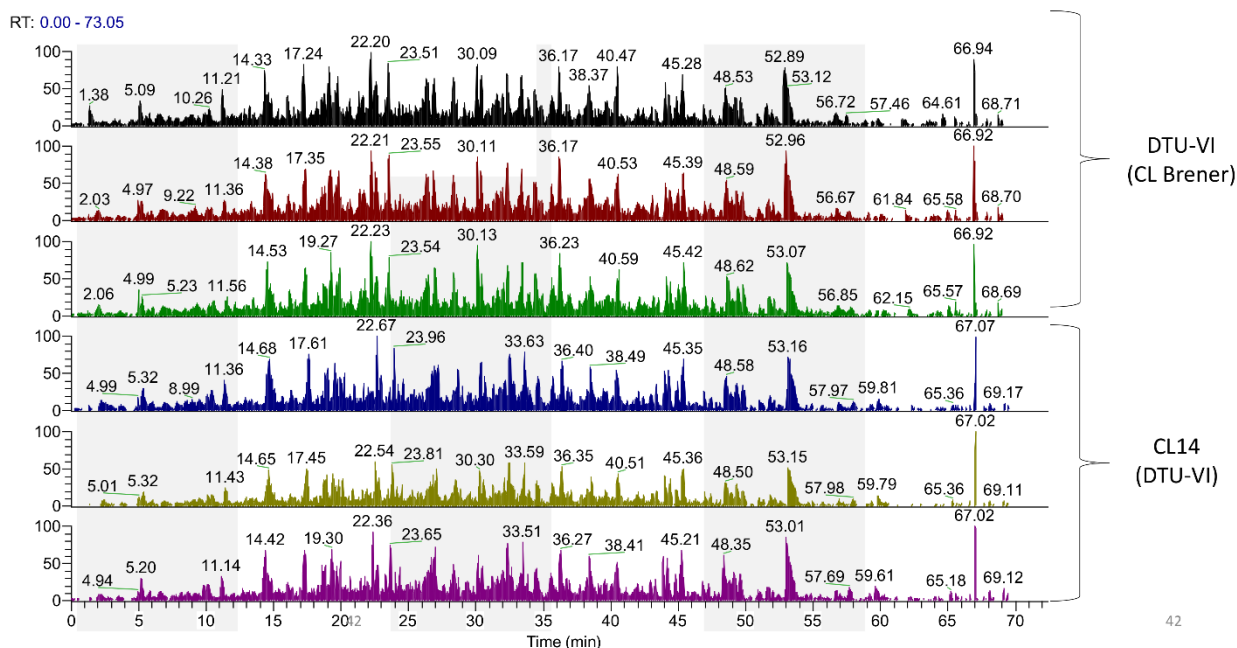
Além disso, o desempenho da abordagem de correspondência espectral Tc-STAMS<sup>2</sup> foi testado para identificar corretamente os conjuntos de dados MS/MS de: 1) uma cepa de *T. cruzi* que é conhecida por pertencer a DTU-VI (*CL14*) (35), 2) de uma espécie filogeneticamente relacionada, como *T. vivax* e 3) de espécies com genoma completamente distantes (ex: humano, *E. coli* e camundongo). Para esta análise, foi construída uma nova biblioteca MS/MS utilizando os espectros de todas as seis DTUs, incluindo *T. cruzi CL14* e *T. vivax* na forma metacíclica (meta1 e meta2). Primeiramente, os espectros MS/MS da cepa *T. cruzi CL14* foram comparadas com a biblioteca e o *score* de similaridade correspondente ao *CL14* foi de 0.417. Curiosamente, embora os *scores* de similaridade de *CL14* com DTU-I e V foram próximas de zero, a semelhança entre *CL14* e DTU-VI foi comparativamente alta (*score* = 0.133), indicando que muitos espectros MS/MS de *CL14* são compartilhados com a DTU-VI (*CL Brener*), como mostra a **Figura 15**.



**Figura 15** – Matrix de validação do método Tc-STAMS<sup>2</sup>. Foi construída uma biblioteca com o score do SpectraST das seis DTUs, cepa *CL14* e *T. vivax*. A abordagem Tc-STAMS<sup>2</sup> foi testada contra: 1) a cepa *CL14 T. cruzi* (DTU-VI), 2) o conjunto de dados *T. vivax* e 3) conjuntos de dados LC-MS/MS de *E. coli* e tecido humano e camundongo. A sensibilidade da abordagem Tc-STAMS<sup>2</sup> foi testada para a detecção de cepas intra-DTU, tais como a cepa *CL14* e *CL Brener* pertencentes a DTU-VI. Além disso, a especificidade da abordagem Tc-STAMS<sup>2</sup> foi testada para a atribuição de espectros MS/MS derivados de organismos filogeneticamente distantes, como camundongos e humanos. Em particular, a biblioteca espectral MS/MS usando sete cepas pertencentes as seis DTUs, tais como, *Sylvio X10 cl1* (DTU-I), *Y* (DTU-II), *M6241 cl6* (DTU-III), *CanIII cl1* (DTU-IV), *MN cl2* (DTU-V), *CL Brener* e *CL14* (DTU-VI) e dados MS/MS de *T. vivax* (formas epimastigota, metacíclica e sanguínea) foram adicionadas a biblioteca espectral. Corridas de LC-MS/MS independentes das diferentes DTUs, fases de vida de *T. vivax*, tecido de camundongo, tecido humano e *E. coli* foram comparadas contra a biblioteca espectral MS/MS utilizando o software SpectraST. A discriminação intra-DTU foi conseguida para *CL14* e *CL Brener* e não foi feita nenhuma atribuição para as amostras de *E. coli*, tecido de camundongo e humano. Os espectros MS/MS de *T. vivax* foram atribuídos especificamente a *T. vivax* sem nenhuma identificação de *T. cruzi*.

Deve-se notar que os perfis cromatográficos LC-MS da cepa *CL14* e *CL Brener* tem alta semelhança, (**Figura 16**). No entanto, o Tc-STAMS<sup>2</sup> foi capaz de diferenciar entre as duas cepas dentro da mesma DTU.





**Figura 16** – Perfil cromatográfico LC-MS dos peptídeos pertencentes às cepas *CL Brener* e *CL14* ambas DTU-VI. Três réplicas para cada cepa.

Posteriormente, os espectros MS/MS de três diferentes formas de vida do *T. vivax*, Meta3 – forma metacíclica, BSF1 – forma sanguínea e EP1 – forma epimastigota, foram comparadas contra a biblioteca espectral. Com base no *score* de similaridade, o Tc-STAMS<sup>2</sup> foi capaz de identificar corretamente as amostras de *T. vivax* (**Figura 15**).

Para testar o método com controle negativo, os espectros MS/MS de espécies não relacionadas com o *T. cruzi*, como tecido humano e *E. coli*, foram comparados com a biblioteca. Para essas amostras, os *scores* de similaridade foram próximos de zero, indicando que os *scores* de similaridade encontrados entre dois conjuntos de dados MS/MS são específicos e não aleatórios (**Figura 15**).

Para complementar as análises, foi realizado outro teste no intuito de verificar se mais de uma cepa em uma mesma DTU comprometeria a identificação. Para tal, foi construída outra biblioteca espectral com cepas adicionais, como mostrado na **Tabela 5**, com as cepas *T. cruzi* Sylvio X10/4 e *T. cruzi* *G mucuræ* ambas DTU-I; *T. cruzi* *esmeraldo* DTU-II; *T. cruzi* 3869 e TCC *T. cruzi* (*Trypanosomatid culture collection*, TCC) DTU-III; *T. cruzi* *jose julio* DTU-IV; *T. cruzi* NR CL3 e *T. cruzi* 92:80 CL2 DTU-V; *T. cruzi* *bat* CL1.1 como representante do grupo *Tc bat* (DTU-VII) e outras como *T. marinkellei*, *T. dionisii*, *T. erneyi* e *T. rangeli* AM-80. Como mostra a **Tabela 5**, a metodologia Tc-STAMS<sup>2</sup> foi capaz de identificar cada uma das cepas mesmo algumas mostrando similaridade. A metodologia mostrou-se ser muito sensível, como por exemplo, *T. cruzi* *mucuræ* e *T. cruzi* Sylvio

X10/4 ambas DTU-I mostraram similaridade (0.636 e 0.119 respectivamente). Em alguns casos as similaridades ocorreram com cepas pertencentes a mesma DTU, como as cepas *T. cruzi jose julio* e *T. cruzi camIII CL1* (0.443 e 0.233, respectivamente) ambas cepas são da DTU-IV. Em outros casos ocorreram similaridade entre diferentes DTUs, como no caso das cepas *TCC T. cruzi* (DTU-III), *T. cruzi NR CL3* e *T. cruzi 92:80 CL2* (ambas DTU-V) que mostraram *scores* de similaridade de 0.175, 0.400 e 0.112, respectivamente. Outro caso foi o das cepas *T. cruzi 92:80 CL2* com *T. cruzi NR CL3* (ambas DTU-V) e *T. cruzi CL14* (DTU-VI) que mostraram *scores* de 0.126, 0.381 e 0.140 respectivamente. As similaridades das diferentes DTUs, como as DTUs III e V, e as DTUs V e VI, podem ser explicados pelas suas origens, ambas DTUs V e VI tiveram origem provável entre as DTUs II e III (52-55), portanto, as similaridades entre essas DTUs podem confirmar as suas origens.

Também foi avaliado se misturas de DTUs interferem na identificação. Para tal, utilizamos as cepas *T. cruzi Sylvio X10 CL1* (DTU-I) e *T. cruzi Y* (DTU-II) na proporção 1:1. Os resultados mostraram que houve a identificação da cepa *T. cruzi Sylvio X10 CL1* mostrando *score* de 0.276 contra 0.073 de *scores* para *T. cruzi Y* (**Tabela 5**). Na proporção de 1:0.5 não houve mudanças, a identificação continuou no *T. cruzi Sylvio X10 CL1*, porém com um aumento no *score* (*score* = 0.349). Quando se comparou uma mistura com três cepas diferentes na proporção de 1:1:1, das cepas *T. cruzi Sylvio X10 CL1* (DTU-I), *T. cruzi Y* (DTU-II) e *T. cruzi M6241 CL1* (DTU-III), os *scores* foram 0.161, 0.062 e 0.144, respectivamente. O resultado para *T. cruzi bat CL1.1* (*Tc bat*/DTU-VII) com *score* de 0.887 não mostrou similaridade com nenhuma DTU, fornecendo mais uma indicativa que este grupo é distinto dos outros.

Adicionalmente, foi comparado cepas de outros tripanossomatídeos, como o, *T. marinkellei*, *T. dionisii*, *T. erneyi* e *T. rangeli AM 80* (**Tabela 5**). Os resultados mostraram que não houve nenhuma similaridade com as DTUs, sendo que os *scores* obtidos ficaram por volta de 0.8-0.9, notando que esses tripanossomatídeos estão bem distantes das DTUs. Todas as ocorrências obtiveram um bootstrap de 1.

Tabela 5 - Todas as identificações das DTUs e *T. cruzi-like*.

Cepas	DTU-I			DTU-II		DTU-III			DTU-IV		DTU-V			DTU-VI		Tc bat	Outros			
	<i>Tc Sylvio X10/4</i>	<i>Tc Sylvio X10 CL1</i>	<i>Tc G mucurae</i>	<i>Tc Y</i>	<i>Tc Esmeraldo</i>	<i>Tc 3869</i>	<i>Tc M 6241 CL6</i>	<i>TCC Tc</i>	<i>Tc CanIII CL1</i>	<i>Tc jose julio</i>	<i>Tc MN CL2</i>	<i>Tc NR CL3</i>	<i>Tc 9280 CL2</i>	<i>Tc CL14</i>	<i>Tc CLBrener</i>	<i>Tc bat CL1.1</i>	<i>T marinkellei</i>	<i>T dionisii</i>	<i>Terneyi</i>	<i>T rangeli AM 80</i>
<i>Tc Sylvio X10/4</i>	0.57	0.003	0.065	0	0	0	0.017	0.021	0.012	0.01	0	0.009	0.007	0	0	0.002	0.01	0	0.001	0
<i>Tc Sylvio X10 CL1</i>	0	0.77	0.001	0	0.001	0.014	0	0	0	0	0.022	0	0	0	0.015	0.002	0	0	0.001	0
<i>Tc G mucurae<sup>a</sup></i>	0.119	0.003	0.636	0	0	0	0.022	0.015	0.027	0.011	0	0.017	0.014	0	0	0.003	0.014	0	0.001	0.001
<i>Tc Y</i>	0	0.021	0	0.5	0.002	0.056	0	0	0.036	0	0.017	0	0	0	0.035	0.001	0	0	0.001	0
<i>Tc Esmeraldo</i>	0	0.002	0.001	0	0.314	0.001	0	0	0	0	0.002	0	0	0	0.002	0.043	0	0.004	0.005	0.002
<i>Tc 3869</i>	0	0.013	0.001	0	0.001	0.691	0	0	0	0	0.021	0	0	0	0.014	0.001	0	0	0	0
<i>Tc M 6241 CL6</i>	0.019	0	0.025	0	0	0	0.465	0.024	0.013	0.026	0	0.013	0.015	0.029	0	0.001	0.012	0	0	0
<i>TCC Tc</i>	0.072	0	0.009	0	0.002	0.005	0.033	0.743	0.015	0.01	0.001	0.036	0.021	0.054	0.001	0.013	0.017	0	0.001	0.001
<i>Tc CanIII CL1</i>	0.036	0	0.023	0	0	0	0.041	0.019	0.686	0.085	0	0.015	0.025	0.03	0	0.003	0.019	0	0.001	0.001
<i>Tc jose julio<sup>b</sup></i>	0.047	0	0.035	0	0	0	0.042	0.04	0.233	0.443	0.001	0.023	0.024	0.04	0	0.007	0.032	0	0.001	0.001
<i>Tc MN CL2</i>	0	0.017	0.001	0	0.001	0.021	0	0	0	0	0.752	0	0.002	0	0.016	0.002	0	0	0	0
<i>Tc NR CL3<sup>c</sup></i>	0.059	0.001	0.016	0	0.001	0	0.019	0.175	0.01	0.014	0.002	0.4	0.112	0.076	0	0.003	0.03	0	0	0
<i>Tc 9280 CL2<sup>d</sup></i>	0.072	0.001	0.021	0	0	0	0.04	0.071	0.032	0.014	0.002	0.126	0.381	0.14	0	0.004	0.026	0	0.001	0
<i>Tc CL14</i>	0	0.023	0.001	0	0.003	0.044	0	0	0	0	0.035	0	0	0	0.103	0.001	0	0.001	0.001	0
<i>Tc CLBrener</i>	0.001	0.033	0.001	0	0.002	0.022	0	0.001	0	0	0.039	0	0	0	0.554	0.003	0	0	0	0
<i>Tc bat CL1.1</i>	0	0.001	0	0	0.014	0	0	0	0	0	0	0	0	0	0	0.887	0	0.003	0.003	0.001
<i>T marinkellei</i>	0.008	0	0.005	0	0	0.001	0.008	0.005	0.007	0.005	0	0.007	0.005	0.012	0	0.001	0.874	0.002	0.001	0
<i>T dionisii</i>	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0	0	0.002	0	0.928	0.004	0.002
<i>Terneyi</i>	0	0	0	0	0.002	0	0	0	0	0	0	0	0	0	0	0.002	0	0.004	0.923	0.002
<i>T rangeli AM 80</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.936
DTU-II(1-1) <sup>e</sup>	0	0.276	0.001	0.1	0.005	0.027	0	0	0	0	0.039	0	0	0	0.038	0.011	0	0.001	0.001	0.001
DTU-II(1-0.5) <sup>f</sup>	0	0.349	0	0.1	0.002	0.031	0	0	0	0	0.033	0	0	0	0.039	0.01	0	0.001	0.001	0
DTU-II-III(1-1-1) <sup>g</sup>	0	0.161	0.001	0.1	0.004	0.144	0	0	0	0	0.043	0	0	0	0.037	0.008	0	0.001	0.001	0

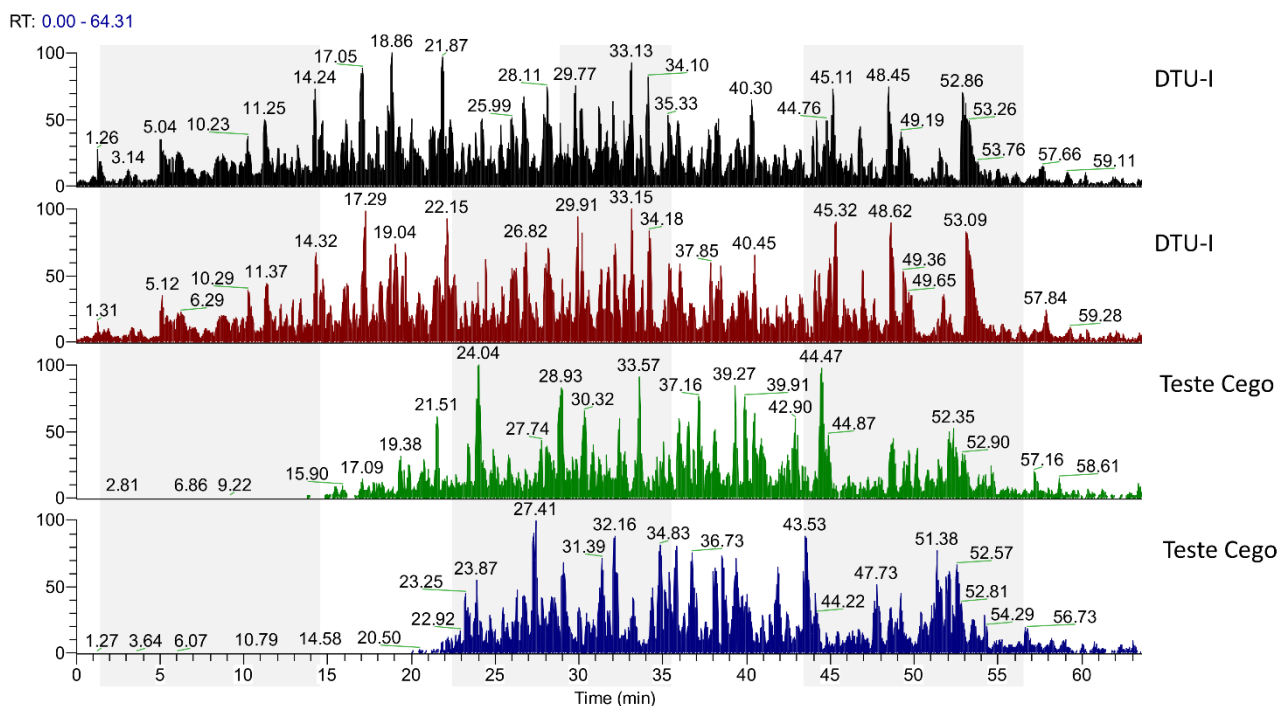
**Tabela 5** – Discriminação de cepas das DTUs de *T. cruzi* e outros tripanossomatídeos baseados em busca de similaridade espectral. Legenda: **a** - *T. cruzi mucurae* contem similaridade com *T. cruzi Sylvio X10/4*; **b** - *T. cruzi jose julio* contem similaridade com *T. cruzi camIII CL1*; **c** - *T. cruzi NR CL3* contem similaridade com *TCC T. cruzi*, *T. cruzi 9280 CL2*; **d** - *T. cruzi 92:80 CL2* contem similaridade com *T. cruzi CL14* e *T. cruzi NR CL3*; **e** - *T. cruzi Sylvio X10 CL1* e *T. cruzi Y*; **f** - *T. cruzi Sylvio X10 CL1* e *T. cruzi Y*; **g** - *T. cruzi Sylvio X10 CL1* com *T. cruzi Y* e *T. cruzi M 6241 CL6*. Os espectros de massas MS/MS das cepas adicionais foram adquiridos pela metodologia HCD no Orbitrap Fusion Lumos.

Amostras	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI	Scores
A1			X				0.139
A2			X				0.137
B	X						0.246
B/ácido 1	X						0.32
B/ácido 2	X						0.309
B/básico 1	X						0.294
B/básico 2	X						0.312
B (20min CID)	X						0.203
B (70min CID)	X						0.267
B (130min CID)	X						0.297
B (70min HCD)	X						0.113
B (20min CID)	X						0.21
B (70min CID)	X						0.279
B (130min CID)	X						0.292
B (70min HCD)	X						0.104

**Figura 17** – Teste do Tc-STAMS<sup>2</sup> quanto a sua robustez para variações técnicas e experimentais. Inicialmente, foram testadas duas comparações interlaboratoriais. Duas DTUs desconhecidas (A e B) foram processadas como descrito na **Figura 12** e adquiridas utilizando o espectrômetro de massas EasynLC acoplado ao LTQ-Orbitrap Velos na *facility* de espectrometria de massas do CEFAP em São Paulo, Brasil. A biblioteca espectral foi construída utilizando a DTU-I (*Sylvio X10 cl1*), DTU-II (*Y*), DTU-III (*M6241 cl6*), DTU-IV (*CanIII cl1*), DTU-V (*MN cl2*), DTU-VI (*CL Brener*) com dados adquiridos no grupo PR em Odense na Dinamarca, usando configurações LC-MS/MS semelhantes (EasynLC acoplado ao LTQ-Orbitrap Velos). As amostras A1 e A2 são uma duplicata biológica da DTU-III (*M6241 cl6*). A amostra B é a DTU-I (*Sylvio X10 cl1*). B/ácido: Os peptídeos derivados da amostra B foram purificados utilizando condições ácidas (0,1% de TFA). B/básico: Os peptídeos derivados da amostra B foram purificados utilizando condições básicas (0,1% de água com amônia). Além disso, diferentes parâmetros analíticos foram alterados a fim de testar a robustez da abordagem Tc-STAMS<sup>2</sup>, como tipo de fragmentação MS/MS, CID (Dissociação induzida por colisão) e HCD (Dissociação induzida por alta energia de colisão). *High* e *Low* indicam a quantidade de amostra injetada no sistema LC-MS, 1 e 0,5 ug, respectivamente. A abordagem foi robusta em relação a diferentes testes analíticos e experimentais.

Além disso, também avaliamos a capacidade dessa estratégia em fornecer a identificação correta de amostras independentes, que foram coletadas e processadas em dias diferentes ou em condições diferentes. Em particular, os diferentes conjuntos de dados foram obtidos em: 1) estudo entre laboratórios, 2) diferentes estratégias de preparação de amostras, 3) diferentes gradientes LC e 4) diferentes métodos MS de fragmentação. Para avaliar a robustez da plataforma Tc-STAMS<sup>2</sup>, outro lote de DTUs foram processados e adquiridos utilizando as condições cromatográficas e de espectrometria de massas semelhantes, conforme descrito na (**Figura 12**), em uma perspectiva de estudo entre laboratórios. Sendo assim, a biblioteca de espectros de massas foi construída com dados adquiridos no grupo PR em Odense, Dinamarca e as amostras de variações técnicas e experimentais foram adquiridas na *facility* de espectrometria de massas do CEFAP em São Paulo. Embora o tipo de instrumento e as condições fossem semelhantes, obteve-se um perfil

cromatográfico diferente (**Figura 18**). No entanto, as duplicatas biológicas desconhecidas A1 e A2 da DTU-III corresponderam corretamente a DTU-III (*score* de similaridade = 0.139) (**Figura 17**). A amostra desconhecida B da DTU-I também apresentou maiores *scores* de similaridade com espectros MS/MS da DTU-I da biblioteca (*score* de similaridade = 0.246). A amostra B foi avaliada em diferentes: 1) condições de preparação, tais como, dessalinização ácida e básica dos peptídeos, 2) diferentes gradientes cromatográficos, tal como, 20, 70 e 130 minutos, 3) diferentes métodos de fragmentação, como CID e HCD com gradiente cromatográfico de 70 minutos e 4) diferentes quantidades de amostra injetada na coluna. Mesmo considerando todas essas fontes técnicas de variações, os *scores* de similaridade continuam a corresponder corretamente à DTU-I mostrando a robustez do Tc-STAMS<sup>2</sup> em diferentes condições experimentais.



**Figura 18** - Perfil cromatográfico LC-MS dos peptídeos pertencentes à cepa *Sylvio X10 cl1* (DTU-I) adquirido no Grupo PR em Odense na Dinamarca e na *facility* de espectrometria de massas do CEFAP na USP, São Paulo, Brasil.

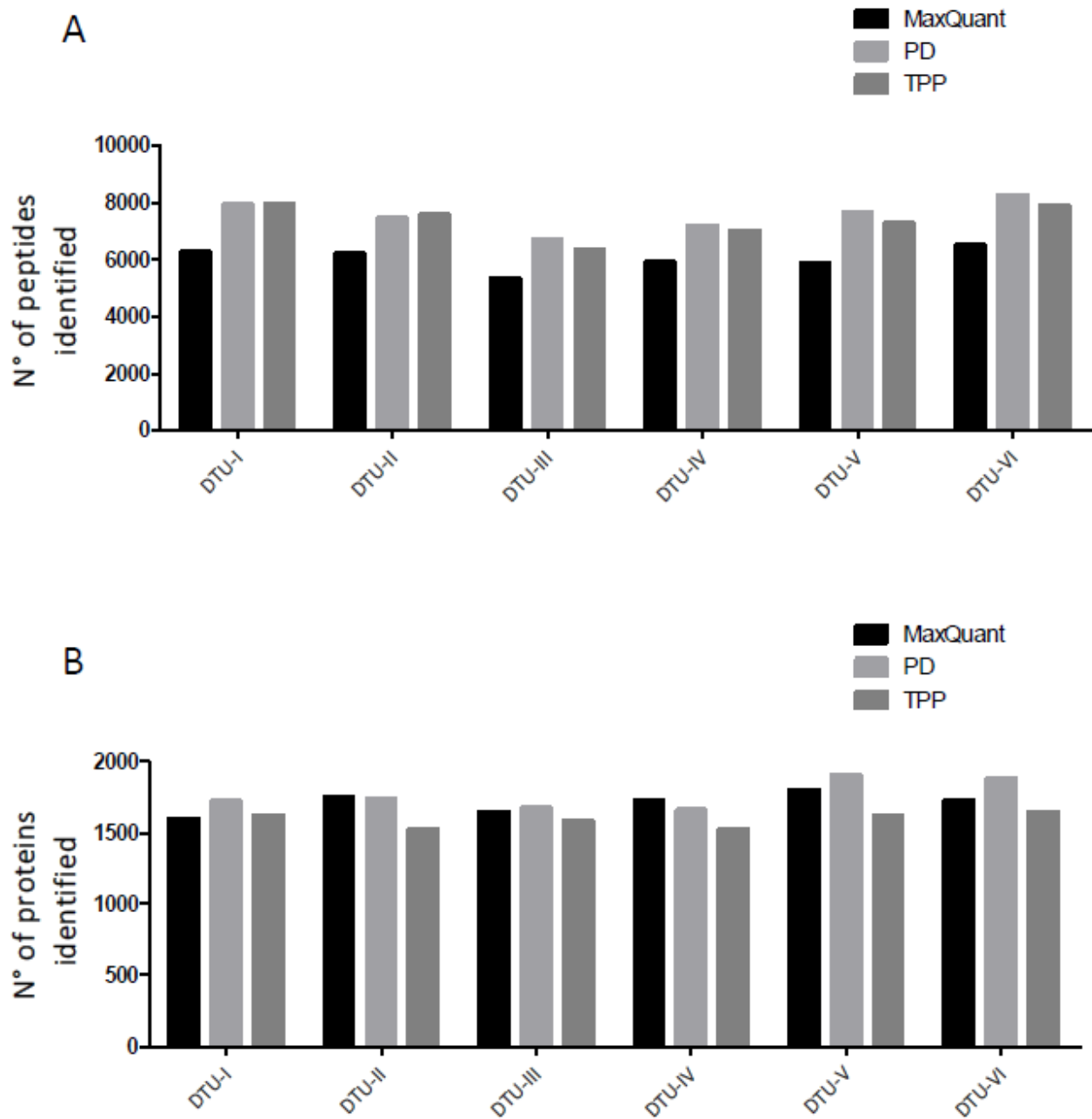
Adicionalmente, foi utilizada outra plataforma de buscas em biblioteca espectral MS/MS, o DiagnoProt, no lugar do SpectraST (194). O DiagnoProt conseguiu diferenciar as diferentes DTUs e associar a cepa *CL14* com o grupo DTU-VI, conforme mostrado no SpectraST, (**Tabela 6**). Tendo em vista esse resultado, dois softwares de busca em biblioteca espectral diferentes poderiam ser implementados no pipeline Tc-STAMS<sup>2</sup> e usados para identificar cepas de *T. cruzi* (**Tabela 3 e 6**).

**Tabela 6** - Correspondência espectral entre as diferentes DTUs utilizando o DiagnoProt. Sete cepas de *T. cruzi* (*Sylvio X10 cl1*, *Y*, *M6241 cl6*, *CanIII cl1*, *MN cl2*, *CL Brener* e *CL14*) pertencentes a seis DTUs foram selecionadas para testar o método Tc-STAMS<sup>2</sup>. O banco de dados foi construído com seis linhagens (*Sylvio X10 cl1*, *Y*, *M6241 cl6*, *CanIII cl1*, *MN cl2* e *CL Brener*) e foi utilizado para comparação das diferentes linhagens, incluindo *CL14*.

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
I	0,174	0,078	0,082	0,089	0,079	0,094
II	0,092	0,172	0,093	0,105	0,088	0,104
III	0,081	0,075	0,162	0,079	0,088	0,092
IV	0,095	0,084	0,082	0,155	0,079	0,089
V	0,070	0,065	0,077	0,063	0,155	0,095
VI	0,088	0,084	0,089	0,077	0,100	0,162
<b>CL14</b>	0,215	0,237	0,246	0,214	0,262	0,360

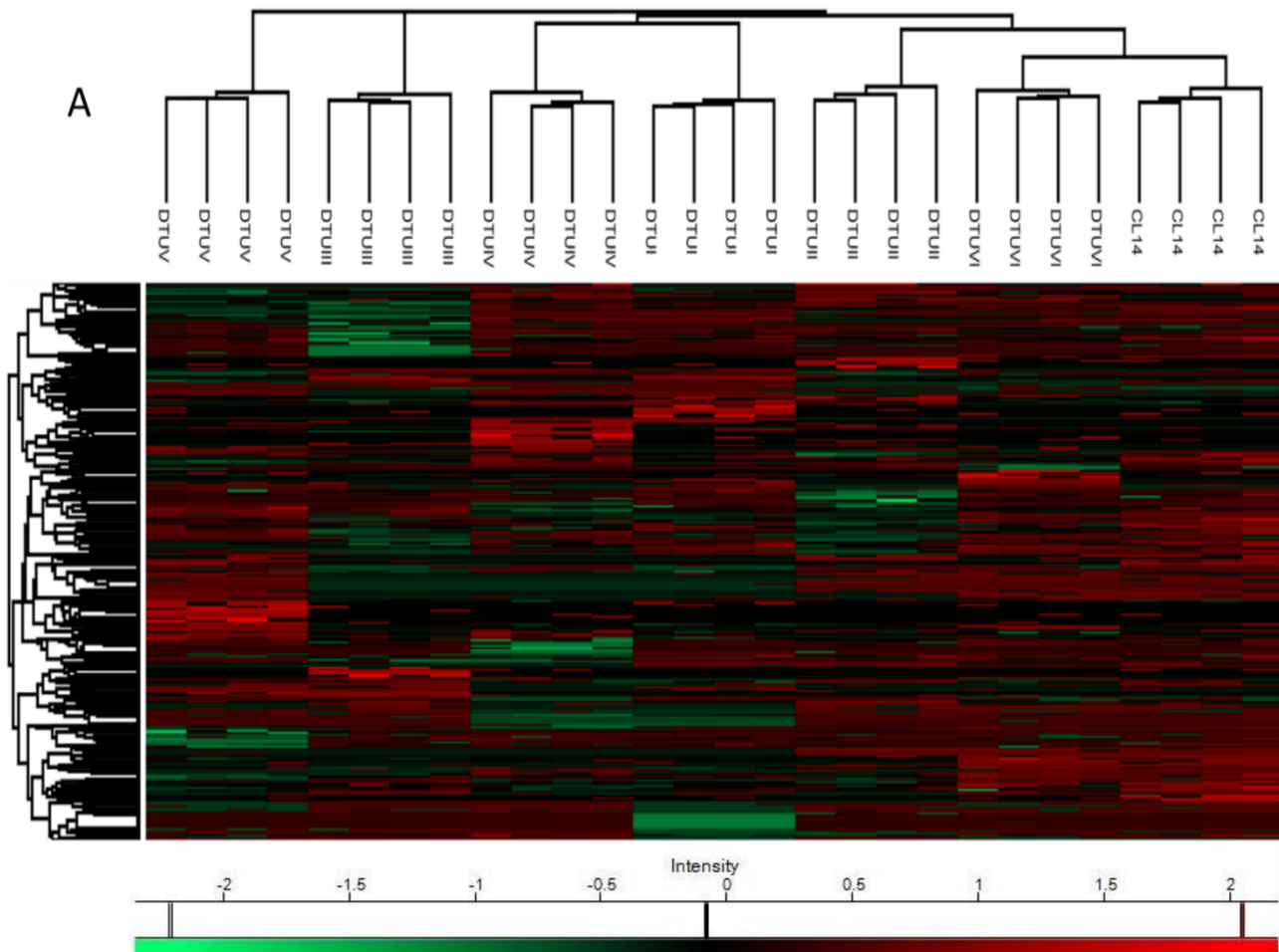
#### 4.2 Análise de *Clustering* Utilizando Identificações de Peptídeos e Proteínas

As buscas em bancos de dados foram realizadas para avaliar a semelhança entre as DTUs usando o resultado das identificações de peptídeos e proteínas. Foram utilizadas três plataformas de buscas em bancos de dados (MaxQuant, TPP e Proteome Discoverer™). A partir de quatro réplicas de cada DTU foram identificados mais de 7000 peptídeos e 4000 proteínas (**Figura 19 e Tabelas Suplementares S1, S2, S3 e S4**). A DTU-I e DTU-VI obtiveram o maior número de identificações devido ao banco de dados de proteínas utilizado para a busca. A cepa *Sylvio* (DTU-I) e *CL Brener* (DTU-VI) são duas cepas de *T. cruzi* cujo genoma foi sequenciado e seu proteoma anotado e depositado no banco de dados Uniprot. Como esperado, apenas 30% do MS/MS totais foram assinados, deixando uma ampla proporção deles não identificados (**Tabela Suplementar S4**).



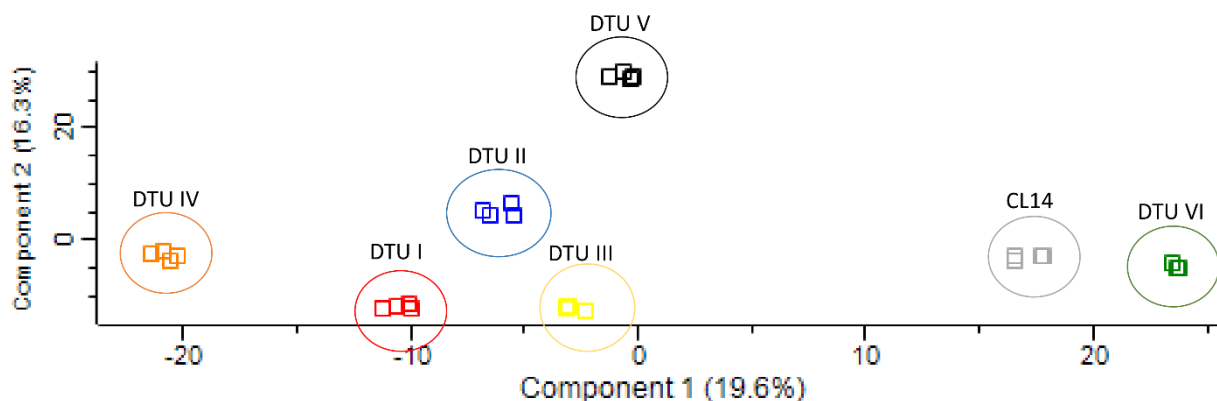
**Figura 19** - A - Identificação dos peptídeos utilizando os softwares MaxQuant, TPP e Proteome Discoverer™ das cepas *Sylvio X10 cl1* (DTU-I), *Y* (DTU-II), *M6241 cl6* (DTU-III), *CanIII cl1* (DTU-IV), *MN cl2* (DTU-V), *CL Brener* (DTU-VI). B – Identificação de proteínas utilizando o MaxQuant, TPP and Proteome Discoverer™ das cepas *Sylvio X10 cl1* (DTU-I), *Y* (DTU-II), *M6241 cl6* (DTU-III), *CanIII cl1* (DTU-IV), *MN cl2* (DTU-V), *CL Brener* (DTU-VI).

A análise de variância (ANOVA  $p < 0.05$  seguido de correção FDR Benjamin-Hochberg) foi aplicada para as proteínas transformadas em  $\log_2$  ou suas intensidades peptídicas previamente identificadas usando o MaxQuant. Assim, 1096 proteínas e 6130 peptídeos apresentaram diferença significativa na abundância entre as seis DTUs (**Tabela Suplementar S5**). Os peptídeos e proteínas diferencialmente expressos foram submetidos à análise de agrupamento (*clustering*) e visualizados em *heat maps*, (**Figura 20A e B**). A cepa *CL14* foi agrupada com *CL Brener* (DTU-VI), indicando alta semelhança no perfil de expressão de proteínas e peptídeos.



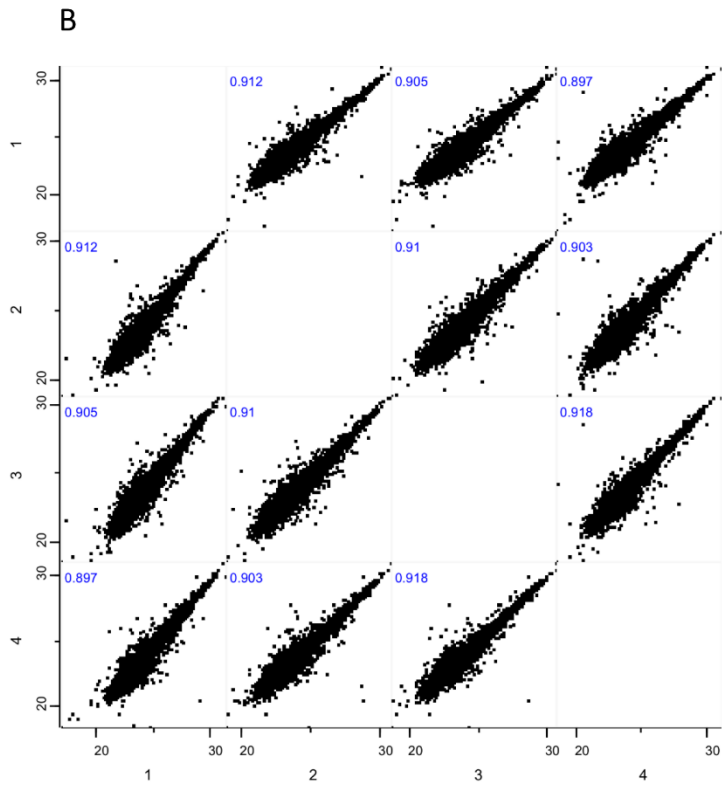
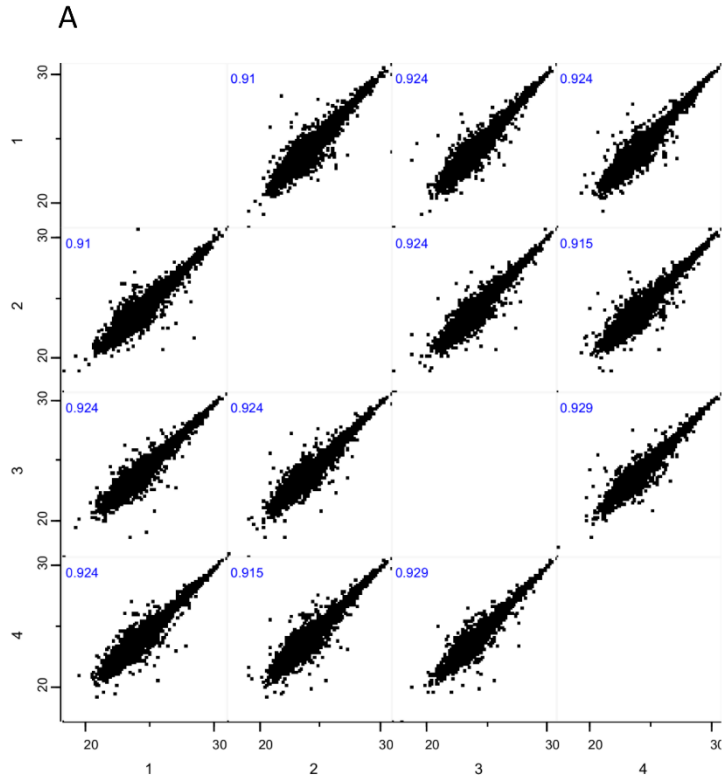


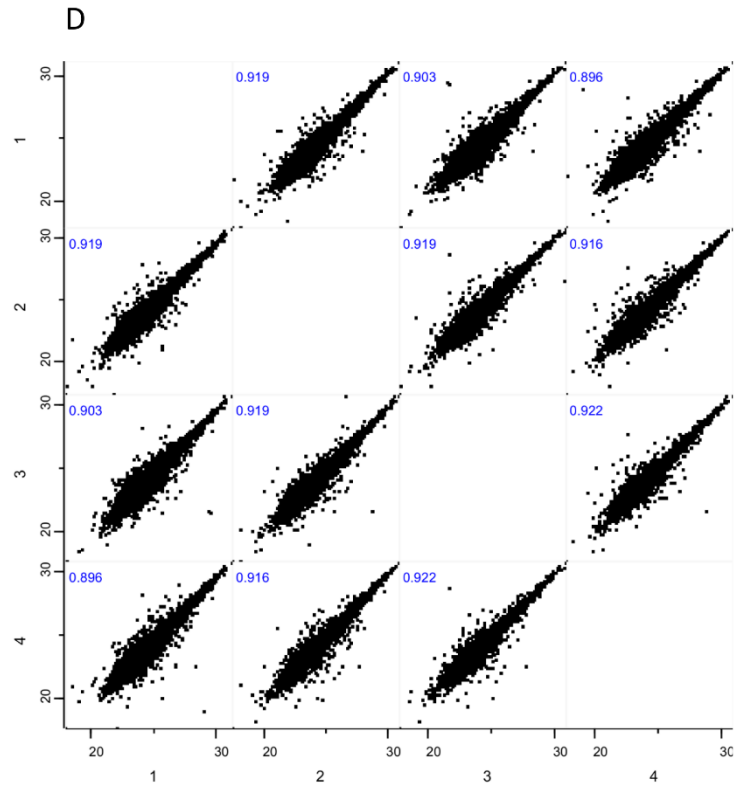
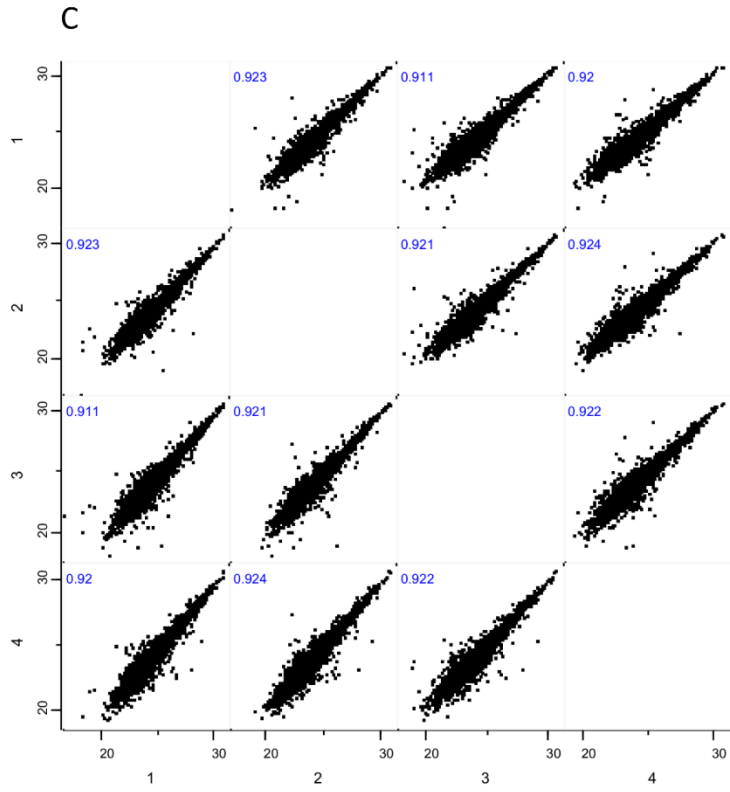


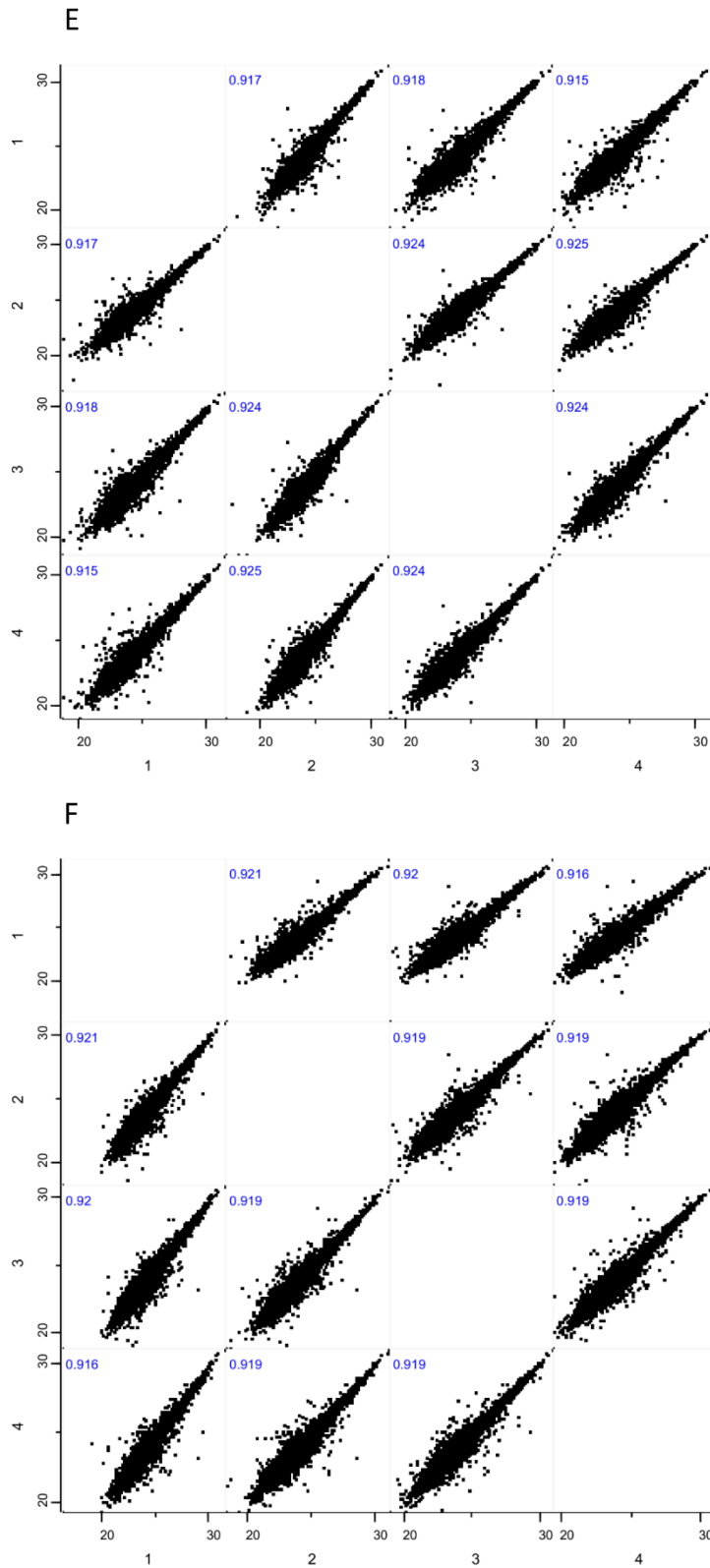


**Figura 21** - Análise de componentes principais (PCA) das proteínas reguladas das setes cepas de *T. cruzi* pertencentes as seis DTUs. Dados LC-MS/MS adquiridos de *Sylvio X10 cl1* (DTU-I), *Y* (DTU-II), *M6241 cl6* (DTU-III), *CanIII cl1* (DTU-IV), *MN cl2* (DTU-V), *CL14* e *CL Brener* (DTU-VI). Busca realizada contra o banco de dados de *T. cruzi* do UniProt (11 de julho de 2017; 51,738 entradas) utilizando a plataforma do MaxQuant como descrito no passo C da **Figura 12**. As proteínas quantificadas foram analisadas pelo teste ANOVA com correção de Benjamin-Hochberg. As 1096 proteínas diferencialmente expressas ( $qvalue < 0,05$ ) foram usadas para construir o PCA. DTU I - Vermelho, DTU II - Azul, DTU III - Amarelo, DTU IV - Laranja, DTU V - Preto, CL Brener - Verde e CL14 - Cinza. Todas as DTUs foram separadas e duas cepas pertencentes ao DTU-VI (CL Brener e CL14) se agruparam.

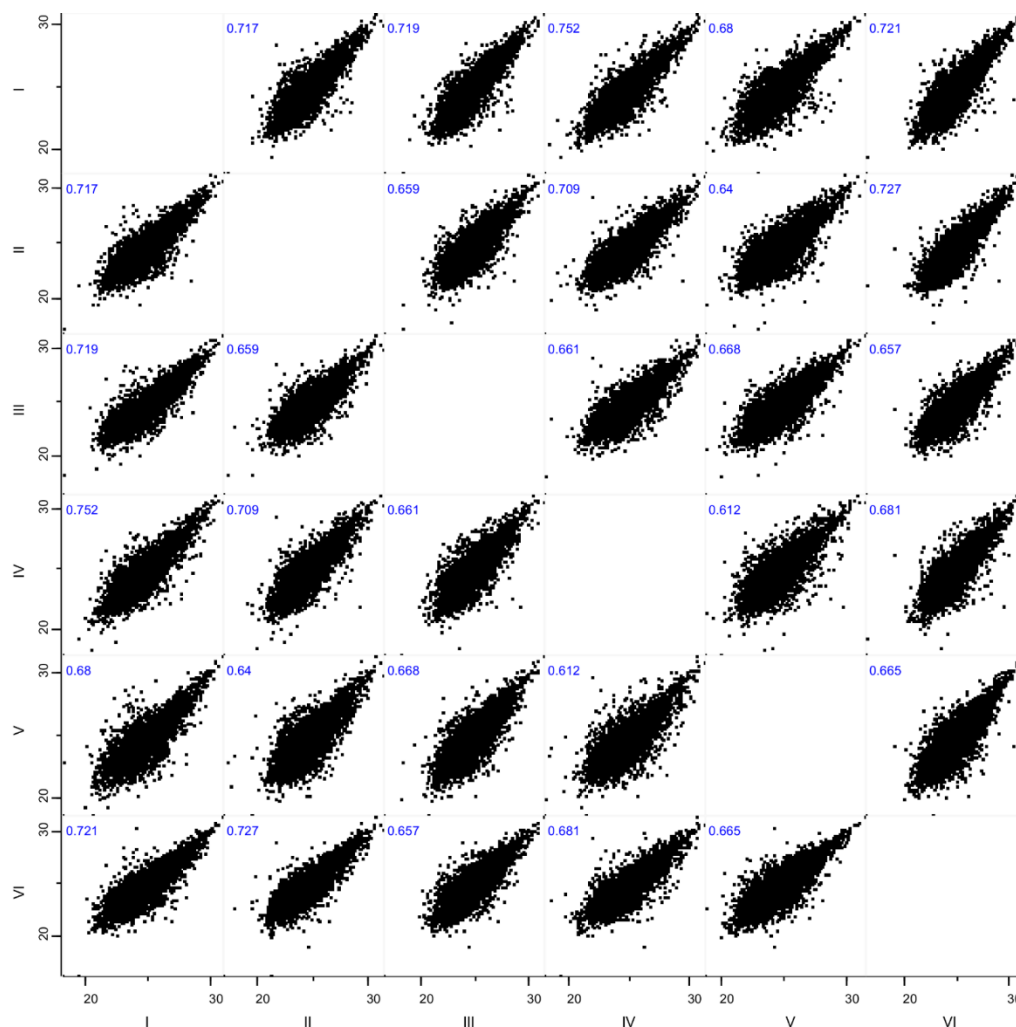
Para verificar a reprodutibilidade dos dados proteômicos, uma análise de correlação ( $R^2$ ) entre as réplicas biológicas de cada DTU e entre as diferentes DTUs foi realizada usando as intensidades normalizadas em  $\log_2$ . Conforme mostrado na (**Figura 22**), valores elevados de correlação foram observados entre as réplicas ( $R^2 > 0.9$ ). Interessantemente, quando comparamos DTUs diferentes (**Figura 23**), o  $R^2$  para 0.6-0.7, mas observou-se uma alta correlação entre *CL14* e *CL Brener* ( $R^2 = 0.83$ ), confirmando uma vez mais a semelhança entre essas duas cepas (**Figura 23**).







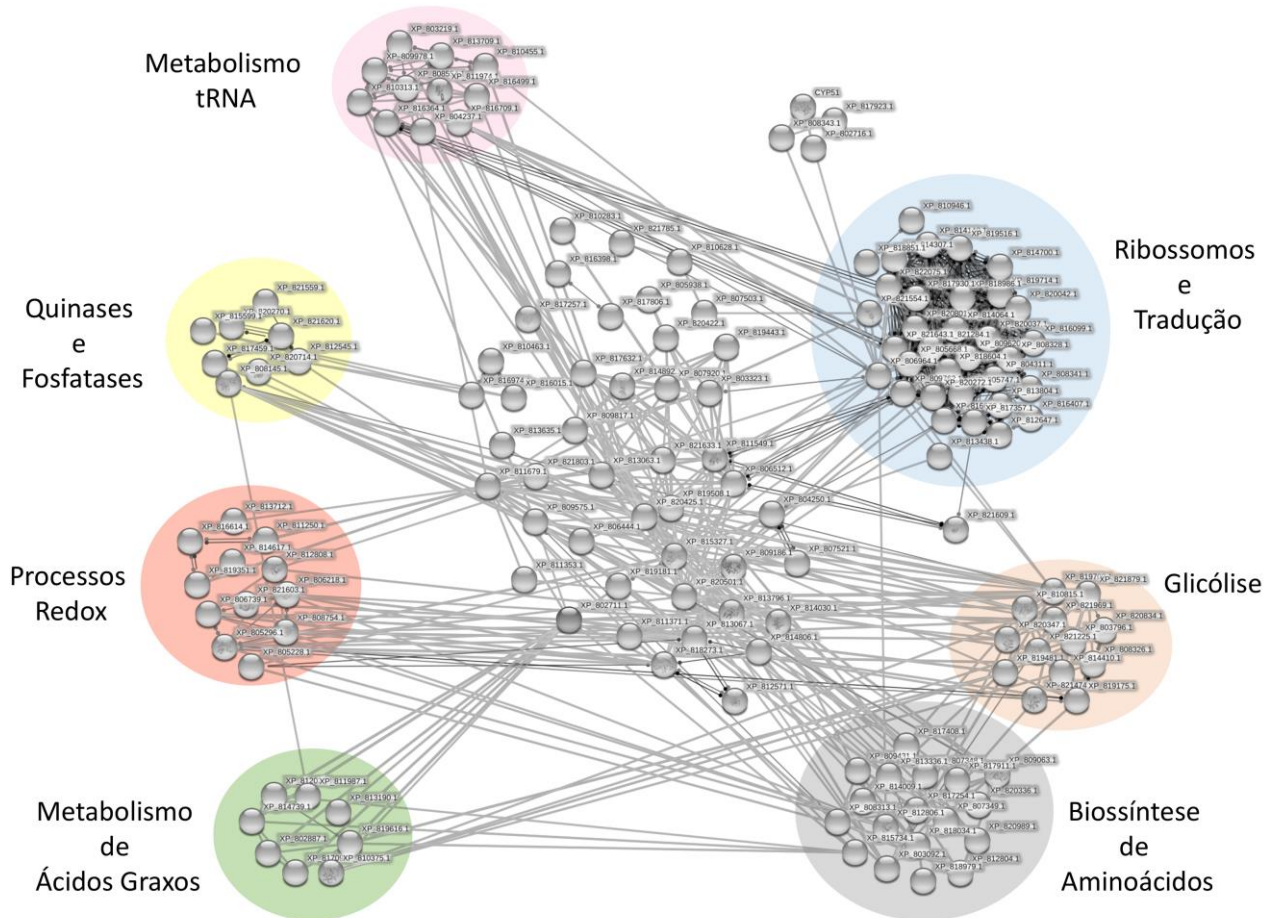
**Figura 22** - *Multi Scatter Plot* obtido comparando as proteínas quantificadas das quatro réplicas das seis DTUs utilizando cada uma das proteínas quantificadas. A - *Sylvio X10 cl1* (DTU-I), B - *Y* (DTU-II), C - *M6241 cl6* (DTU-III), D - *CanIII cl1* (DTU-IV), E - *MN cl2* (DTU-V), F - *CL Brener* (DTU-VI).



**Figura 23** - Multi Scatter Plot obtido comparando os peptídeos quantificados das seis DTUs. A - *Sylvio X10 cl1* (DTU-I), B - *Y* (DTU-II), C - *M6241 cl6* (DTU-III), D - *CanIII cl1* (DTU-IV), E - *MN cl2* (DTU-V), F - *CL Brener* (DTU-VI).

#### 4.3 Análise de Rede Interação Proteína-Proteína

Para analisar quais redes de interação proteína-proteína estão reguladas entre as cepas *CL14* e *CL Brener* que pertencem a DTU-VI, foram utilizadas 306 proteínas estatisticamente significantes (**Tabela Suplementar S6**) (ANOVA  $p < 0.05$  seguido de correção FDR Benjamin-Hochberg). Utilizando o software String (199), os processos mais enriquecidos foram proteínas envolvidas com os ribossomos e tradução, biossíntese de aminoácidos, metabolismo de tRNA, processos *redox*, metabolismo de ácidos graxos, glicólise, quinases e fosfatases, como mostrado na **Figura 24**.



**Figura 24** - Rede de interações proteína-proteína utilizando 306 proteínas construída utilizando o software String. Foram comparadas as cepas *CL14* e *CL Brener* que pertencem a DTU-VI. Os processos mais enriquecidos são: proteínas envolvidas com os ribossomos e tradução, biossíntese de aminoácidos, metabolismo de tRNA, processos redox, metabolismo de ácidos graxos, glicólise, quinases e fosfatases.

## 5 DISCUSSÃO

Muitos estudos empregaram metodologias baseadas em espectrometria de massas para identificar organismos, como bactérias (119). Önder et al. foi o pioneiro no uso de bibliotecas espectrais MS/MS para identificar com precisão de qual animal o carrapato *Ixodes scapularis* se alimentou, mesmo que a alimentação tenha ocorrido meses antes (173). Outras estratégias combinaram a informação genética em conjunto com a identificação de peptídeos baseado em espectrometria de massas para a correta identificação de microorganismos (201-203). Mais recentemente, Shao et. al. (174) demonstrou que é possível identificar cepas de *E. coli* só com o espectro de fragmentação MS/MS sem a necessidade de identificação dos peptídeos (174).

No presente estudo, descrevemos uma metodologia de correspondência espectral MS/MS, independente de genoma, rápido, preciso e projetada para identificar diferentes DTUs de *T. cruzi*, denominada de Tc-STAMS<sup>2</sup>. Em primeiro lugar, foi construída uma biblioteca espectral de peptídeos MS/MS usando três réplicas de cada uma das seis DTUs. A quarta réplica foi utilizada para testar a capacidade do método para diferenciar cada DTUs usando a biblioteca de referência. Conforme mostrado na **Tabela 3, 4 e Figura 21** essa abordagem foi capaz de diferenciar DTUs usando o software SpectraST. Foi realizada uma atribuição única para a DTU correta. O método foi testado usando um conjunto de dados de espectros de peptídeos MS/MS obtido a partir de diferentes condições de crescimento de *T. cruzi*.

O próximo passo foi testar a biblioteca espectral MS/MS em relação a organismos filogeneticamente relacionados, como *T. vivax* e organismos distantes como *E. coli* e *homo sapiens* (**Figura 15**). Para tal teste, foi construído outro banco de dados com os mesmos espectros MS/MS das DTUs e da cepa *CL14* e *T. vivax*. Claro que, mesmo com milhares de espectros de fragmentação, os *scores* obtidos ao comparar amostras de organismos como o humano ou *E. coli* com a biblioteca foram próximos de zero, demonstrando a especificidade do método na identificação de amostras da espécie/cepa cujo espectro MS/MS está presente na biblioteca.

Também mostramos que o método de correspondência espectral é robusto, mesmo com fontes de variações inter e intra-laboratoriais usando o mesmo ou diferentes espectrômetros de massas. Além do mais, mesmo com dados obtidos de diferentes laboratórios, mudança na preparação da amostra, na cromatografia e no método de fragmentação, ainda foi possível identificar corretamente amostras de DTU-III e DTU-I, como mostrado na **Figura 17**.



E para complementar as análises anteriores, foi realizada uma segunda comparação, com mais representantes de cada DTU e também com outros tripanossomatídeos. Como foi mostrado na **Tabela 5**, a metodologia Tc-STAMS<sup>2</sup> foi capaz de distinguir cada DTU como também verificar similaridade entre cepas da mesma DTU e entre DTUs. Em relação a similaridade entre DTUs, como no caso das DTUs III, V e VI podem estar relacionadas a sua origem. Essas DTUs são resultado de hibridização entre as DTUs II e III (36, 52-55), portanto esse resultado pode ser utilizado como mais uma prova para confirmar sua origem de hibridização. Também verificamos se a metodologia era capaz de distinguir co-infecção. No primeiro momento foi simulado uma co-infecção em um microtúbulo, isto é, misturamos duas DTUs, a DTU-I (*Sylvio X10 cl1*) e DTU-II (*Y*), na proporção 1:1. O resultado foi positivo para DTU-I (**Tabela 5**). Depois queríamos verificar se a mistura entre DTU-I e DTU-II com proporções de 1:0.5 poderiam influenciar na identificação. Como mostra a **Tabela 5**, houve identificação somente da DTU-I. E por fim, verificamos se com três DTUs com proporções de 1:1:1 a metodologia Tc-STAMS<sup>2</sup> teria capacidade de identificar cada uma delas. Para tal, utilizamos as cepas *Sylvio X10 cl1* (DTU-I), *Y* (DTU-II) e *M6241 cl6* (DTU-III). Foi identificado a DTU-I e DTU-III não identificando a DTU-II. A metodologia mostrou-se ser incapaz na identificação de co-infecção, tendo tendência de identificar apenas uma DTU (**Tabela 5**). O que pode ter influenciado os resultados podem ter relação com a preparação das amostras, como fatores 1) manuais, na hora de misturar as DTUs podem ter ido mais células de uma DTU do que de outra, 2) se tivéssemos misturado peptídeos ao invés de células poderia ter sido mais uniforme a quantidade de cada DTU ou 3) na hora da pipetagem pode ter ocorrido algum erro e ter ido uma maior quantidade de uma certa DTU do que outra.

Adicionalmente, comparamos outros tripanossomatídeos, como o *T. cruzi bat CL1.1*, *T. marinkellei*, *T. dionisii*, *T. erneyi* e *T. rangeli AM 80*, como mostra a **Tabela 5**, os resultados foram específicos para cada tripanossomatídeo, indicando que mesmo sendo morfológicamente indistinguíveis, esses tripanossomatídeos apresentam proteínas diferentes resultando em marcadores moleculares distintos, características bioquímicas, imunológicas e fenotípicas que os caracterizam como *T. cruzi-like* (204-212).

Curiosamente, nós observamos que o *score* de similaridade é dependente do número de espectros MS/MS adquiridos. Quanto maior o tempo de gradiente utilizado, maior a separação de peptídeos antes da análise de MS e maior o número de espectros adquiridos, resultando em maiores

scores para a mesma amostra em comparação com a biblioteca. O uso de diferentes métodos de fragmentação também mostrou que (CID) fornece pontuações maiores que HCD em 70 min de tempo de gradiente. Embora o HCD forneça alta resolução de espectros MS/MS, CID fornece um sequenciamento MS/MS mais rápido, gerando assim mais espectros que podem ser combinados com a biblioteca.i

Os métodos atuais utilizados para identificar DTUs são difíceis de implementar exigindo pessoal muito bem treinado (59, 61). Outros métodos requerem o conhecimento do genoma para identificar as DTUs e muitos dos métodos não podem diferenciar cepas com genoma similar. Nossa plataforma é projetada para complementar as metodologias já desenvolvidas e para auxiliar nas identificações das DTUs ou qualquer outro parasita. Esta plataforma é fácil de implementar, rápida e requer apenas o espectro MS/MS para identificação.

Além disso, os espectros de fragmentação também foram submetidos à análise de busca em banco de dados para identificação de peptídeos e proteínas usando os softwares MaxQuant, Trans-Proteomic Pipeline (TPP) e Proteome Discoverer. Em geral, o número de identificações de proteínas e peptídeos foi reprodutível e consistente entre diferentes softwares de busca de banco de dados. O maior número de proteínas e peptídeos foi observado para DTU-I e VI. Embora o número de proteínas anotadas no banco de dados UniProt seja maior para DTUs I e VI, o número de proteínas identificadas para DTUs V, III e IV não foi significativamente menor. Devido a isso, uma abordagem proteômica baseada em espectrometria de massas pode ser usada para comparar quantitativamente a expressão proteica de diferentes DTUs, mesmo com diferenças na anotação do genoma entre elas, e usar essas informações para identificar vias específicas das DTUs e correlacioná-las com o fenótipo da cepa.

Além do mais, foi possível agrupar a DTU-VI e *CL14* usando as proteínas diferencialmente expressas ou os peptídeos obtidos pelo teste de ANOVA (**Figuras 20 e 21**). Este resultado valida o dado que o *CL14* pertence a DTU-VI (35), e também validou os resultados obtidos com a correspondência espectral, onde o *score* de similaridade entre *CL14* e DTU-VI foi maior em comparação com as outras DTUs. Além disso, a análise de PCA também foi capaz de discriminar as seis DTUs e determinar a semelhança entre a *CL14* e DTU-VI, como observado pela proximidade entre esses dois grupos.

A análise de *clustering* utilizando a identificação das proteínas em diferentes DTUs também foi realizada por Telleria et, al. (179), no entanto, apenas 261 (experimento 1) e 172 (experimento 2) dos *spot* dos géis 2-DE foram considerados para esta análise. Em nosso estudo, a análise de *clustering* foi realizada usando 1096 proteínas diferencialmente expressas entre DTUs, aumentando o número de características para construir um agrupamento robusto do *T. cruzi* (**Figura 20**).

Os resultados quantitativos foram utilizados para estudar as diferenças de sinalização molecular entre as várias cepas. Em nossa análise de interação proteína-proteína, escolhemos as cepas *CL14* e *CL Brener* que pertencem a DTU-VI. A cepa *CL Brener* é uma cepa de referência que teve sua origem de isolado de *Triatoma infestans* (213) tem preferencialmente tropismo para células musculares do coração e esquelética (214). Já a cepa *CL14* é uma cepa de baixa virulência (215) comparado com sua cepa de origem (cepa CL) (216). Essas cepas estão na mesma DTU (DTU-VI), porém com características bem diferentes. Na análise de interação proteína-proteína entre as cepas (**Figura 24**), foram identificados processos específicos, alguns deles relacionados com a virulência, por exemplo, as proteínas envolvidas nos processos de resistência ao estresse oxidativo como a Tripanotiona redutase (*Trypanothione reductase* TR, processos *redox* **Figura 24**) (217) estão *up*-regulada em *CL Brener* em relação ao *CL14*. redução da tripanotiona disulfeto (T[S]2) em tripanotiona ditiol (T[SH]2), desencadeando, assim, uma cascata de eventos responsáveis pela neutralização de espécies reativas de oxigênio. Desta forma, a TR mantém um ambiente redutor no interior do parasito, protegendo-o contra o estresse oxidativo. A TR é uma molécula antioxidante muito importante em tripanossomatídeos e foi proposta como alvo terapêutico (217). O nível de tripanotiona redutase foi identificado aumentado durante a diferenciação da forma epimastigota para tripomastigota metacíclico, sugerindo uma correlação com o processo de virulência. As cepas (por exemplo CL Brener) mais virulentas mostraram uma *up*-regulação dessa proteína (218, 219).

Entre as proteínas reguladas encontramos as calpainas, proteinases cálcio dependentes envolvidas em vários processos como apoptose, transdução de sinais e remodelamento do citoesqueleto (220). As nossas análises revelaram que a calpaina está relacionada com a proliferação do parasito e a resistência a fármacos (221). Adicionalmente, encontramos diferentemente regulada a enzima bifuncional diidrofolato redutase-timidase sintase (DHFR-TS), responsável pela replicação celular. A perda dessa enzima resulta em parasitos atenuados, como mostrado por Perez Brandan C, et al., (222). Todos esses processos estão *up*-regulados em *CL Brener* (**Tabela Suplementar S6**). Essa

metodologia demonstrou a capacidade de identificar fatores de virulência, e estudos com um número maior de cepas poderiam identificar novos alvos de virulência.

A estratégia Tc-STAMS<sup>2</sup> é robusta, precisa, rápida, fácil de realizar e completamente automatizada. Trabalhos de Dworzanski, J. P. (203) e Shao W. et al., (174) usaram uma metodologia similar para a identificação de bactérias, no entanto, esta é a primeira vez que a combinação espectral é aplicada para discriminar diferentes DTUs de *T. cruzi*.

Recentemente, foi desenvolvido em nosso laboratório uma estratégia baseada em MALDI-TOF para a identificação direta de tripanossomatídeos com base no perfil MS, denominado de Identificação Direta de Tripanossomatídeos por Ionização por Dissorção a Laser Assistida por Matriz por Tempo de Vôo (DIT-MALDI TOF) (223). No entanto, esse método não tem resolução para discriminar DTUs usando o perfil MS. Diferentemente do Tc-STAMS<sup>2</sup> que é capaz de discriminar cepas de uma mesma DTUs e de DTUs diferentes (**Tabelas 3 e 5**).

## 6 CONCLUSÃO

Neste estudo, apresentamos uma caracterização de DTUs de *T. cruzi* com base na correspondência espectral MS/MS. O método Tc-STAMS<sup>2</sup> é robusto, sensível e poderoso com base na identificação pelo espectro MS/MS mostrando ser um método de caracterização para complementar outros já estabelecidos. O método proposto pode ajudar pesquisadores na caracterização das DTUs sem a necessidade de dados genômicos, o que é muito escasso para esse parasito.

## REFERÊNCIAS\*

\*De acordo com:

International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to biomedical journals. [2011 Jul 15]. Available from: [http://www.nlm.nih.gov/bsd/uniform\\_requirements.html](http://www.nlm.nih.gov/bsd/uniform_requirements.html)

1. Chagas C. Nova tripanozomíase humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. Memórias do Instituto Oswaldo Cruz. 1909;1:159-218.
2. WHO. Chagas disease (American trypanosomiasis): World Health Organization 2017 [Available from: <http://www.who.int/mediacentre/factsheets/fs340/en/>].
3. PAHO. Chagas disease. Pan American Health Organization; 2017.
4. Rassi A, Jr., Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010;375(9723):1388-402.
5. Victora CG, Barreto ML, do Carmo Leal M, Monteiro CA, Schmidt MI, Paim J, et al. Health conditions and health-policy innovations in Brazil: the way forward. Lancet. 2011;377(9782):2042-53.
6. Martins-Melo FR, Ramos AN, Jr., Alencar CH, Heukelbach J. Prevalence of Chagas disease in Brazil: a systematic review and meta-analysis. Acta Trop. 2014;130:167-74.
7. Hotez PJ, Fujiwara RT. Brazil's neglected tropical diseases: an overview and a report card. Microbes Infect. 2014;16(8):601-6.
8. Dias JC, Ramos AN, Jr., Gontijo ED, Luquetti A, Shikanai-Yasuda MA, Coura JR, et al. [Brazilian Consensus on Chagas Disease, 2015]. Epidemiol Serv Saude. 2016;25(spe):7-86.
9. Dias JCP, Coura JR, Yasuda MAS. The present situation, challenges, and perspectives regarding the production and utilization of effective drugs against human Chagas disease. Revista da Sociedade Brasileira de Medicina Tropical. 2014;47:123-5.
10. Dias J, Silveira A, Schofield C. The impact of Chagas disease control in Latin America: a review. Memórias do Instituto Oswaldo Cruz. 2002;97:603-12.
11. Barreto ML, Teixeira MG, Bastos FI, Ximenes RA, Barata RB, Rodrigues LC. Successes and failures in the control of infectious diseases in Brazil: social and environmental context, policies, interventions, and research needs. Lancet. 2011;377(9780):1877-89.
12. Martins-Melo FR, Ramos Junior AN, Alencar CH, Heukelbach J. Multiple causes of death related to Chagas' disease in Brazil, 1999 to 2007. Revista da Sociedade Brasileira de Medicina Tropical. 2012;45:591-6.
13. Martins-Melo FR, Ramos AN, Jr., Alencar CH, Heukelbach J. Mortality due to Chagas disease in Brazil from 1979 to 2009: trends and regional differences. J Infect Dev Ctries. 2012;6(11):817-24.
14. da Nobrega AA, de Araujo WN, Vasconcelos AM. Mortality due to Chagas disease in Brazil according to a specific cause. Am J Trop Med Hyg. 2014;91(3):528-33.
15. Saúde Md. Boletim Epidemiológico. Ministério da Saúde - Secretaria de Vigilância em Saúde 2015.
16. Cavalier-Smith T. Eukaryote kingdoms: Seven or nine? Biosystems. 1981;14(3):461-81.
17. Vickerman K. The diversity of the kinetoplastid flagellates. In Biology of the Kinetoplastida. Evans WHRLDA, editor: Academic Press; 1976.

18. Liu B, Liu Y, Motyka SA, Agbo EE, Englund PT. Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol.* 2005;21(8):363-9.
19. Moreira D, Lopez-Garcia P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol.* 2004;54(Pt 5):1861-75.
20. Simpson AGB, Roger AJ. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Molecular Phylogenetics and Evolution.* 2004;30(1):201-12.
21. Jankevicius JV, Itow-Jankevicius S, Maeda LA, Campaner M, Conchon I, do Carmo JB, et al. [Biological cycle of *Phytomonas*]. *Mem Inst Oswaldo Cruz.* 1988;83 Suppl 1:601-10.
22. Leonard G, Soanes DM, Stevens JR. Resolving the question of trypanosome monophyly: a comparative genomics approach using whole genome data sets with low taxon sampling. *Infect Genet Evol.* 2011;11(5):955-9.
23. Maslov DA, Votypka J, Yurchenko V, Lukes J. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol.* 2013;29(1):43-52.
24. Lukes J, Skalicky T, Tyc J, Votypka J, Yurchenko V. Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol.* 2014;195(2):115-22.
25. Porcel BM, Deneud F, Opperdoes F, Noel B, Madoui MA, Hammarton TC, et al. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet.* 2014;10(2):e1004007.
26. Levine ND. **The Trypanosomes of Mammals.** A Zoological Monograph. Cecil A. Hoare. Blackwell, Oxford, England, 1972 (U.S. distributor, Davis, Philadelphia). xviii, 750 pp. + plates. \$34.50. *Science.* 1973;179(4068):60-.
27. A Podlipaev S, Frolov A. [The phylogeny of the Trypanosomatidae: the molecular and morphological approaches]2000. 169-82 p.
28. Bonaldo MC, Souto-Padron T, de Souza W, Goldenberg S. Cell-substrate adhesion during *Trypanosoma cruzi* differentiation. *J Cell Biol.* 1988;106(4):1349-58.
29. Kleffmann T, Schmidt J, Schaub GA. Attachment of *Trypanosoma cruzi* epimastigotes to hydrophobic substrates and use of this property to separate stages and promote metacyclogenesis. *J Eukaryot Microbiol.* 1998;45(5):548-55.
30. Tyler KM, Engman DM. The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol.* 2001;31(5-6):472-81.
31. Brener Z. *Trypanosoma cruzi*: morfologia e ciclo evolutivo. In: João Carlos Pinto Dias JRC, editor. *Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral.* Scielo Books: FIOCRUZ; 1997. p. 486.
32. Rodriguez A, Webster P, Ortego J, Andrews NW. Lysosomes behave as Ca<sup>2+</sup>-regulated exocytic vesicles in fibroblasts and epithelial cells. *J Cell Biol.* 1997;137(1):93-104.
33. Andrews NW. Living dangerously: how *Trypanosoma cruzi* uses lysosomes to get inside host cells, and then escapes into the cytoplasm. *Biol Res.* 1993;26(1-2):65-7.
34. CDC. Parasites - American Trypanosomiasis (also known as Chagas Disease) Center for Disease Control and Prevention2017 [Available from: <https://www.cdc.gov/parasites/chagas/biology.html>].
35. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz.* 2009;104(7):1051-4.

36. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol.* 2012;12(2):240-53.
37. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology.* 2009;136(6):641-55.
38. Lima L, Espinosa-Alvarez O, Ortiz PA, Trejo-Varon JA, Carranza JC, Pinto CM, et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Trop.* 2015;151:166-77.
39. Andrade SG. Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano. *Revista Patologia Tropical.* 31974. p. 65-121.
40. Magalhães JB, Andrade SG. Estudo do comportamento de cepas de *Trypanosoma cruzi* após passagem em diferentes espécies de triatomíneos. *Revista da Sociedade Brasileira de Medicina Tropical.* 1991;24:209-16.
41. Filardi LS, Brener Z. A rapid method for testing in vivo the susceptibility of different strains of *Trypanosoma cruzi* to active chemotherapeutic agents. *Mem Inst Oswaldo Cruz.* 1984;79(2):221-5.
42. Miles MA, Toye PJ, Oswald SC, Godfrey DG. The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Trans R Soc Trop Med Hyg.* 1977;71(3):217-25.
43. Tibayrenc M, Ayala FJ. Isozyme Variability in *Trypanosoma Cruzi*, the Agent of Chagas' Disease: Genetical, Taxonomical, and Epidemiological Significance. *Evolution.* 1988;42(2):277-92.
44. Miles MA, Souza A, Pova M, Shaw JJ, Lainson R, Toye PJ. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature.* 1978;272(5656):819-21.
45. Miles MA, Lanham SM, de Souza AA, Pova M. Further enzymic characters of *Trypanosoma cruzi* and their evaluation for strain identification. *Trans R Soc Trop Med Hyg.* 1980;74(2):221-37.
46. Tibayrenc M, Ward P, Moya A, Ayala FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proc Natl Acad Sci U S A.* 1986;83(1):115-9.
47. Tibayrenc M, Cariou ML, Solignac M, Carlier Y. Genetic arguments against actual mating in *Trypanosoma cruzi*. Taxonomic implications. *Comptes Rendus des Seances de l'Academie des Sciences - Series III.* 1981;293(3):207-9.
48. Steindel M, Dias Neto E, de Menezes CL, Romanha AJ, Simpson AJ. Random amplified polymorphic DNA analysis of *Trypanosoma cruzi* strains. *Mol Biochem Parasitol.* 1993;60(1):71-9.
49. Morel C, Chiari E, Camargo EP, Mattei DM, Romanha AJ, Simpson L. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proc Natl Acad Sci U S A.* 1980;77(11):6810-4.
50. Henriksson J, Aslund L, Macina RA, Franke de Cazzulo BM, Cazzulo JJ, Frasch AC, et al. Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and



- further demonstration of karyotype variability in *Trypanosoma cruzi*. *Mol Biochem Parasitol*. 1990;42(2):213-23.
51. Macedo AM, Melo MN, Gomes RF, Pena SD. DNA fingerprints: a tool for identification and determination of the relationships between species and strains of *Leishmania*. *Mol Biochem Parasitol*. 1992;53(1-2):63-70.
  52. Brisse S, Barnabe C, Banuls AL, Sidibe I, Noel S, Tibayrenc M. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. *Mol Biochem Parasitol*. 1998;92(2):253-63.
  53. Sturm NR, Vargas NS, Westenberger SJ, Zingales B, Campbell DA. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *Int J Parasitol*. 2003;33(3):269-79.
  54. Westenberger SJ, Barnabe C, Campbell DA, Sturm NR. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics*. 2005;171(2):527-43.
  55. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *Int J Parasitol*. 2009;39(12):1305-17.
  56. Ochaya S. Comparative genomics and molecular characterization of N-alpha Acetyltransferase in Trypanosomes for drug target identification: Karolinska Institutet; 2013.
  57. Macedo AM, Machado CR, Oliveira RP, Pena SD. *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease. *Mem Inst Oswaldo Cruz*. 2004;99(1):1-12.
  58. Zingales B. *Trypanosoma cruzi*: um parasita, dois parasitas ou vários parasitas da doença de chagas? 2011. 44-8 p.
  59. Cosentino RO, Aguero F. A simple strain typing assay for *Trypanosoma cruzi*: discrimination of major evolutionary lineages from a single amplification product. *PLoS Negl Trop Dis*. 2012;6(7):e1777.
  60. Tibayrenc M, Neubauer K, Barnabe C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proc Natl Acad Sci U S A*. 1993;90(4):1335-9.
  61. Brisse S, Verhoef J, Tibayrenc M. Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *Int J Parasitol*. 2001;31(11):1218-26.
  62. Cura CI, Duffy T, Lucero RH, Bisio M, Peneau J, Jimenez-Coello M, et al. Multiplex Real-Time PCR Assay Using TaqMan Probes for the Identification of *Trypanosoma cruzi* DTUs in Biological and Clinical Samples. *PLoS Negl Trop Dis*. 2015;9(5):e0003765.
  63. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94(3):441-8.
  64. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 2005;309(5733):409-15.
  65. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Negl Trop Dis*. 2011;5(3):e984.

66. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, Segovia M, et al. Trypanosoma cruzi Ilc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. *PLoS Negl Trop Dis*. 2009;3(9):e510.
67. Grisard EC, Teixeira SM, de Almeida LG, Stoco PH, Gerber AL, Talavera-Lopez C, et al. Trypanosoma cruzi Clone Dm28c Draft Genome Sequence. *Genome Announc*. 2014;2(1).
68. Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, Lewis MD, et al. Comparative genomic analysis of human infective Trypanosoma cruzi lineages with the bat-restricted subspecies T. cruzi marinkellei. *BMC Genomics*. 2012;13:531.
69. Bartholomeu DC, de Paiva RM, Mendes TA, DaRocha WD, Teixeira SM. Unveiling the intracellular survival gene kit of trypanosomatid parasites. *PLoS Pathog*. 2014;10(12):e1004399.
70. Kramer S, Carrington M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol*. 2011;27(1):23-30.
71. Johnson PJ, Kooter JM, Borst P. Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell*. 1987;51(2):273-81.
72. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell*. 2003;11(5):1291-9.
73. De Gaudenzi JG, Noe G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. *Essays Biochem*. 2011;51:31-46.
74. Agabian N. Trans splicing of nuclear pre-mRNAs. *Cell*. 1990;61(7):1157-60.
75. Preusser C, Jae N, Bindereif A. mRNA splicing in trypanosomes. *Int J Med Microbiol*. 2012;302(4-5):221-4.
76. Stuart K, Allen TE, Heidmann S, Seiwert SD. RNA editing in kinetoplastid protozoa. *Microbiol Mol Biol Rev*. 1997;61(1):105-20.
77. Hajduk S, Ochsenreiter T. RNA editing in kinetoplastids. *RNA Biol*. 2010;7(2):229-36.
78. Rudenko G, Chung HM, Pham VP, Van der Ploeg LH. RNA polymerase I can mediate expression of CAT and neo protein-coding genes in Trypanosoma brucei. *EMBO J*. 1991;10(11):3387-97.
79. Lee MG, Van der Ploeg LH. Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol*. 1997;51:463-89.
80. Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walte. *Molecular Biology of the Cell*: Garland Science; 2002.
81. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev*. 2007;21(15):1833-56.
82. Zomerdijk JC, Kieft R, Borst P. Efficient production of functional mRNA mediated by RNA polymerase I in Trypanosoma brucei. *Nature*. 1991;353(6346):772-5.
83. Rudenko G, Blundell PA, Dirks-Mulder A, Kieft R, Borst P. A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in T. brucei. *Cell*. 1995;83(4):547-53.
84. Yan S, Lodes MJ, Fox M, Myler PJ, Stuart K. Characterization of the Leishmania donovani ribosomal RNA promoter. *Mol Biochem Parasitol*. 1999;103(2):197-210.
85. McAndrew M, Graham S, Hartmann C, Clayton C. Testing Promoter Activity in the Trypanosome Genome: Isolation of a Metacyclic-Type VSG Promoter, and Unexpected Insights into RNA Polymerase II Transcription. *Experimental Parasitology*. 1998;90(1):65-76.

86. Parsons M, Nelson RG, Watkins KP, Agabian N. Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell*. 1984;38(1):309-16.
87. Araujo PR, Teixeira SM. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. *Mem Inst Oswaldo Cruz*. 2011;106(3):257-66.
88. Dias JC, Coura J. Clínica e Terapêutica da Doença de Chagas: uma abordagem prática para o clínico geral. João Carlos Pinto Dias JRC, editor. FIOCRUZ: FIOCRUZ; 1997.
89. Antas PR, Medrano-Mercado N, Torrico F, Ugarte-Fernandez R, Gomez F, Correa Oliveira R, et al. Early, intermediate, and late acute stages in Chagas' disease: a study combining anti-galactose IgG, specific serodiagnosis, and polymerase chain reaction analysis. *Am J Trop Med Hyg*. 1999;61(2):308-14.
90. Andrade DV, Gollob KJ, Dutra WO. Acute chagas disease: new global challenges for an old neglected disease. *PLoS Negl Trop Dis*. 2014;8(7):e3010.
91. Pudukollu G, Gowda RM, Khan IA, Navarro VS, Vasavada BC. Clinical aspects of the Chagas' heart disease. *Int J Cardiol*. 2007;115(3):279-83.
92. Linardi DP, NALDMPM. Parasitologia Humana: Atheneu; 2005.
93. Prata A. Clinical and epidemiological aspects of Chagas disease. *Lancet Infect Dis*. 2001;1(2):92-100.
94. Docampo R. Sensitivity of parasites to free radical damage by antiparasitic drugs. *Chem Biol Interact*. 1990;73(1):1-27.
95. Rodrigues Coura J, de Castro SL. A critical review on Chagas disease chemotherapy. *Mem Inst Oswaldo Cruz*. 2002;97(1):3-24.
96. Bern C, Montgomery SP, Herwaldt BL, Rassi A, Jr., Marin-Neto JA, Dantas RO, et al. Evaluation and treatment of chagas disease in the United States: a systematic review. *JAMA*. 2007;298(18):2171-81.
97. Gomes ML, Galvao LM, Macedo AM, Pena SD, Chiari E. Chagas' disease diagnosis: comparative analysis of parasitologic, molecular, and serologic methods. *Am J Trop Med Hyg*. 1999;60(2):205-10.
98. Carlier Y, Torrico F, Sosa-Estani S, Russomando G, Luquetti A, Freilij H, et al. Congenital Chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women. *PLoS Negl Trop Dis*. 2011;5(10):e1250.
99. Moreira LM. Ciências Genômicas: Fundamentos e Aplicações: Editora Cubo; 2015.
100. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*. 1995;16(7):1090-4.
101. Yates JR, 3rd. Mass spectrometry and the age of the proteome. *J Mass Spectrom*. 1998;33(1):1-19.
102. de Hoog CL, Mann M. Proteomics. *Annu Rev Genomics Hum Genet*. 2004;5:267-93.
103. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
104. Bensimon A, Heck AJ, Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem*. 2012;81:379-405.
105. Anderson NL, Hofmann JP, Gemmell A, Taylor J. Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis. *Clin Chem*. 1984;30(12 Pt 1):2031-6.

106. Tarroux P, Vincens P, Rabilloud T. HERMeS: A second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part V: Data analysis. *ELECTROPHORESIS*. 1987.
107. Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics*. 2009;2009.
108. Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*. 2001;73(23):5683-90.
109. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*. 1999;17(7):676-82.
110. Rabilloud T, Chevallet M, Luche S, Lelong C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics*. 2010;73(11):2064-77.
111. Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF. Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. *Electrophoresis*. 1998;19(8-9):1501-5.
112. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis*. 2000;21(6):1104-15.
113. Rabilloud T. Membrane proteins and proteomics: love is possible, but so difficult. *Electrophoresis*. 2009;30 Suppl 1:S174-80.
114. Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. *J Chromatogr A*. 2010;1217(25):3938-45.
115. Hilton GR, Benesch JL. Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J R Soc Interface*. 2012;9(70):801-16.
116. Mallick P, Kuster B. Proteomics: a pragmatic perspective. *Nat Biotechnol*. 2010;28(7):695-709.
117. Barbosa EB, Vidotto A, Polachini GM, Henrique T, Trovó de Marqui AB, Tajara EH. Proteômica: metodologias e aplicações no estudo de doenças humanas. *Revista da Associação Médica Brasileira*. 2012;58(3):366-75.
118. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*. 2013;14(1):35-48.
119. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat Genet*. 2003;33 Suppl:311-23.
120. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol*. 2004;5(9):699-711.
121. Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom*. 2000;35(7):804-17.
122. Carr SA, Huddleston MJ, Bean MF. Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci*. 1993;2(2):183-96.
123. Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov*. 2003;2(2):140-50.
124. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 2006;312(5771):212-7.
125. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207.
126. Lanças FM. A cromatografia líquida moderna e a espectrometria de massas: Finalmente "compatíveis"? II. A escolha do analisador de massas. *Scientia Chromatographica*. 2013.

127. Das A. Spinning Charged Test-Particles in General Relativity. *Progress of Theoretical Physics*. 1960;23(4):610-5.
128. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*. 2000;72(6):1156-62.
129. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. *J Mass Spectrom*. 2005;40(4):430-43.
130. Dass C. Mass Analysis and Ion Detection. *Fundamentals of Contemporary Mass Spectrometry*: John Wiley & Sons, Inc.; 2006. p. 67-117.
131. Hart-Smith G, Blanksby SJ. Mass Analysis. *Mass Spectrometry in Polymer Chemistry*: Wiley-VCH Verlag GmbH & Co. KGaA; 2011. p. 5-32.
132. Cantú MD, Carrilho E, Wulff NA, Palma MS. Seqüenciamento de peptídeos usando espectrometria de massas: um guia prático. *Química Nova*. 2008;31:669-75.
133. Hunt DF, Yates JR, 3rd, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A*. 1986;83(17):6233-7.
134. Bythell BJ, Suhai S, Somogyi A, Paizs B. Proton-driven amide bond-cleavage pathways of gas-phase peptide ions lacking mobile protons. *J Am Chem Soc*. 2009;131(39):14057-65.
135. Mann M, Meng CK, Fenn JB. Interpreting mass spectra of multiply charged ions. *Analytical Chemistry*. 1989;61(15):1702-8.
136. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*. 1984;11(11):601.
137. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 2004;101(26):9528-33.
138. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*. 2004;15(1):12-6.
139. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007;4(9):709-12.
140. Cass Q, Cassiano N. *Cromatografia Líquida: Novas Tendências e Aplicações*: Elsevier Academic; 2015.
141. Zhang Y, Ficarro SB, Li S, Marto JA. Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom*. 2009;20(8):1425-34.
142. Kocher T, Pichler P, Schutzbier M, Stingl C, Kaul A, Teucher N, et al. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J Proteome Res*. 2009;8(10):4743-52.
143. Mischerikow N, van Nierop P, Li KW, Bernstein HG, Smit AB, Heck AJ, et al. Gaining efficiency by parallel quantification and identification of iTRAQ-labeled peptides using HCD and decision tree guided CID/ETD on an LTQ Orbitrap. *Analyst*. 2010;135(10):2643-52.
144. Zhang J, Wang Y, Li S. Deuterium isobaric amine-reactive tags for quantitative proteomics. *Anal Chem*. 2010;82(18):7588-95.
145. Johnson JV, Yost RA, Kelley PE, Bradford DC. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Analytical Chemistry*. 1990;62(20):2162-72.
146. Nogueira FC, Domont GB. Survey of shotgun proteomics. *Methods Mol Biol*. 2014;1156:3-23.
147. Mazzarino M, de la Torre X, Di Santo R, Fiacco I, Rosi F, Botre F. Mass spectrometric characterization of tamoxifene metabolites in human urine utilizing different scan

- parameters on liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2010;24(6):749-60.
148. Sadygov RG, Cociorva D, Yates JR, 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods*. 2004;1(3):195-202.
  149. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994;5(11):976-89.
  150. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-805.
  151. Patterson SD. Data analysis--the Achilles heel of proteomics. *Nat Biotechnol*. 2003;21(3):221-2.
  152. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*. 2007;7(5):655-67.
  153. Heller SR. Conversational mass spectral retrieval system and its use as an aid in structure determination. *Analytical Chemistry*. 1972;44(12):1951-61.
  154. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-72.
  155. Yates JR, 3rd, Morgan SF, Gatlin CL, Griffin PR, Eng JK. Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal Chem*. 1998;70(17):3557-65.
  156. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem*. 2006;78(16):5678-84.
  157. Shao W, Lam H. Tandem mass spectral libraries of peptides and their roles in proteomics research. *Mass Spectrom Rev*. 2017;36(5):634-48.
  158. Craig R, Cortens JC, Fenyo D, Beavis RC. Using Annotated Peptide Mass Spectrum Libraries for Protein Identification. *Journal of Proteome Research*. 2006;5(8):1843-9.
  159. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol*. 2004;22(11):1459-66.
  160. Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol*. 2005;6(1):R9.
  161. Stanislaus R, Arthur JM, Rajagopalan B, Moerschell R, McGlothlen B, Almeida JS. An open-source representation for 2-DE-centric proteomics and support infrastructure for data storage and analysis. *BMC Bioinformatics*. 2008;9:4.
  162. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res*. 2004;3(6):1234-42.
  163. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, et al. PRIDE: the proteomics identifications database. *Proteomics*. 2005;5(13):3537-45.
  164. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res*. 2006;34(Database issue):D655-8.

165. Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol.* 2005;6(7):577-83.
166. Lam H, Deutsch EW, Edes JS, Eng JK, Stein SE, Aebersold R. Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods.* 2008;5(10):873-5.
167. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol.* 2005;1:2005 0017.
168. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry.* 1994;5(9):859-66.
169. Pedrioli PG. Trans-proteomic pipeline: a pipeline for proteomic analysis. *Methods Mol Biol.* 2010;604:213-38.
170. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics.* 2010;73(11):2092-123.
171. Lam H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics.* 2011;10(12):R111 008565.
172. Lam H, Aebersold R. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods.* 2011;54(4):424-31.
173. Onder O, Shao W, Kemps BD, Lam H, Brisson D. Identifying sources of tick blood meals using unidentified tandem mass spectral libraries. *Nat Commun.* 2013;4:1746.
174. Shao W, Zhang M, Lam H, Lau SC. A peptide identification-free, genome sequence-independent shotgun proteomics workflow for strain-level bacterial differentiation. *Sci Rep.* 2015;5:14337.
175. Paba J, Ricart CA, Fontes W, Santana JM, Teixeira AR, Marchese J, et al. Proteomic analysis of *Trypanosoma cruzi* developmental stages using isotope-coded affinity tag reagents. *J Proteome Res.* 2004;3(3):517-24.
176. Paba J, Santana JM, Teixeira AR, Fontes W, Sousa MV, Ricart CA. Proteomic analysis of the human pathogen *Trypanosoma cruzi*. *Proteomics.* 2004;4(4):1052-9.
177. Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, et al. The *Trypanosoma cruzi* proteome. *Science.* 2005;309(5733):473-6.
178. Parodi-Talice A, Monteiro-Goes V, Arrambide N, Avila AR, Duran R, Correa A, et al. Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. *J Mass Spectrom.* 2007;42(11):1422-32.
179. Telleria J, Biron DG, Brizard JP, Demetree E, Seveno M, Barnabe C, et al. Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*. *Proc Natl Acad Sci U S A.* 2010;107(47):20411-6.
180. Diaz ML, Torres R, Gonzalez CI. [Differential protein expression in developmental stages of *Trypanosoma cruzi* I isolated from a patient with chronic chagasic cardiomyopathy]. *Biomedica.* 2011;31(4):503-13.
181. Queiroz RM, Charneau S, Motta FN, Santana JM, Roepstorff P, Ricart CA. Comprehensive proteomic analysis of *Trypanosoma cruzi* epimastigote cell surface proteins by two complementary methods. *J Proteome Res.* 2013;12(7):3255-63.
182. Queiroz RM, Charneau S, Bastos IM, Santana JM, Sousa MV, Roepstorff P, et al. Cell surface proteome analysis of human-hosted *Trypanosoma cruzi* life stages. *J Proteome Res.* 2014;13(8):3530-41.

183. Ferella M, Nilsson D, Darban H, Rodrigues C, Bontempi EJ, Docampo R, et al. Proteomics in *Trypanosoma cruzi*--localization of novel proteins to various organelles. *Proteomics*. 2008;8(13):2735-49.
184. Scott DA, Docampo R. Characterization of isolated acidocalcisomes of *Trypanosoma cruzi*. *J Biol Chem*. 2000;275(31):24215-21.
185. dos Santos Junior Ade C, Kalume DE, Camargo R, Gomez-Mendoza DP, Correa JR, Charneau S, et al. Unveiling the *Trypanosoma cruzi* Nuclear Proteome. *PLoS One*. 2015;10(9):e0138667.
186. Batista JA, Teixeira SM, Donelson JE, Kirchhoff LV, de Sa CM. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Mol Biochem Parasitol*. 1994;67(2):301-12.
187. Rout MP, Field MC. Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. *J Biol Chem*. 2001;276(41):38261-71.
188. Miles MA, Cedillos RA, Povoá MM, de Souza AA, Prata A, Macedo V. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? *Lancet*. 1981;1(8234):1338-40.
189. Camargo EP. Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of Metacyclic Trypanosomes in Liquid Media. *Rev Inst Med Trop Sao Paulo*. 1964;6:93-100.
190. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*. 2007;2(8):1896-906.
191. Palmisano G, Parker BL, Engholm-Keller K, Lendal SE, Kulej K, Schulz M, et al. A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Mol Cell Proteomics*. 2012;11(11):1191-202.
192. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008;24(21):2534-6.
193. Efron B, Robert. *An Introduction to the Bootstrap* (Chapman & Hall/CRC Monographs on Statistics & Applied Probability): Chapman and Hall/CRC; 1994.
194. Silva ARF, Lima DB, Leyva A, Duran R, Batthyany C, Aquino PF, et al. DiagnoProt: a tool for discovery of new molecules by mass spectrometry. *Bioinformatics*. 2017;33(12):1883-5.
195. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics*. 2013;13(1):22-4.
196. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002;74(20):5383-92.
197. Lee HJ, Jeong SK, Na K, Lee MJ, Lee SH, Lim JS, et al. Comprehensive genome-wide proteomic analysis of human placental tissue for the Chromosome-Centric Human Proteome Project. *J Proteome Res*. 2013;12(6):2458-66.
198. Jackson AP, Goyard S, Xia D, Foth BJ, Sanders M, Wastling JM, et al. Global Gene Expression Profiling through the Complete Life Cycle of *Trypanosoma vivax*. *PLoS Negl Trop Dis*. 2015;9(8):e0003975.
199. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447-52.



200. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731-40.
201. Jabbour RE, Deshpande SV, Wade MM, Stanford MF, Wick CH, Zulich AW, et al. Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. *Appl Environ Microbiol*. 2010;76(11):3637-44.
202. Dworzanski JP, Snyder AP, Chen R, Zhang H, Wishart D, Li L. Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring. *Anal Chem*. 2004;76(8):2355-66.
203. Dworzanski JP, Deshpande SV, Chen R, Jabbour RE, Snyder AP, Wick CH, et al. Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *J Proteome Res*. 2006;5(1):76-87.
204. Baker JR, Miles MA, Godfrey DG, Barrett TV. Biochemical characterization of some species of *Trypanosoma* (*Schizotrypanum*) from bats (*Microchiroptera*). *Am J Trop Med Hyg*. 1978;27(3):483-91.
205. Ebert F. Comparison of isoenzymes of some species of the subgenus *schizotrypanum* from bats by isoelectrofocusing. *Tropenmed Parasitol*. 1983;34(2):93-7.
206. Tibayrenc M, Le Ray D. General classification of the isoenzymic strains of *Trypanosoma* (*Schizotrypanum*) *cruzi* and comparison with *T. (S.) C. marinkellei* and *T. (Herpetosoma) rangeli*. *Ann Soc Belg Med Trop*. 1984;64(3):239-48.
207. Petry K, Baltz T, Schottelius J. Differentiation of *Trypanosoma cruzi*, *T. cruzi marinkellei*, *T. dionisii* and *T. vespertilionis* by monoclonal antibodies. *Acta Trop*. 1986;43(1):5-13.
208. Petry K, Voisin P, Baltz T. Complex lipids as common antigens to *Trypanosoma cruzi*, *T. dionisii*, *T. vespertilionis* and nervous tissue (astrocytes, neurons). *Acta Trop*. 1987;44(4):381-6.
209. Stevens JR, Noyes HA, Schofield CJ, Gibson W. The molecular evolution of *Trypanosomatidae*. *Adv Parasitol*. 2001;48:1-56.
210. Barnabe C, Brisse S, Tibayrenc M. Phylogenetic diversity of bat trypanosomes of subgenus *Schizotrypanum* based on multilocus enzyme electrophoresis, random amplified polymorphic DNA, and cytochrome b nucleotide sequence analyses. *Infect Genet Evol*. 2003;2(3):201-8.
211. Telleria J, Barnabe C, Hide M, Banuls AL, Tibayrenc M. Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi*. *Mol Biochem Parasitol*. 2004;137(1):133-41.
212. Lima L, Ortiz PA, da Silva FM, Alves JM, Serrano MG, Cortez AP, et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in *Trypanosoma cruzi*, *T. cruzi*-like and other trypanosome species. *PLoS One*. 2012;7(6):e38385.
213. Brener Z, Chiari E. Variações Morfológicas Observadas em Diferentes Amostras de *Trypanosoma cruzi*. *Rev Inst Med Trop Sao Paulo*. 1963.
214. Melo RC, Brener Z. Tissue tropism of different *Trypanosoma cruzi* strains. *J Parasitol*. 1978;64(3):475-82.
215. Atayde VD, Neira I, Cortez M, Ferreira D, Freymuller E, Yoshida N. Molecular basis of non-virulence of *Trypanosoma cruzi* clone CL-14. *Int J Parasitol*. 2004;34(7):851-60.
216. Yoshida N. Surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi*. *Infect Immun*. 1983;40(2):836-9.
217. Fairlamb AH, Cerami A. Metabolism and functions of trypanothione in the Kinetoplastida. *Annu Rev Microbiol*. 1992;46:695-729.

218. Piacenza L, Zago MP, Peluffo G, Alvarez MN, Basombrio MA, Radi R. Enzymes of the antioxidant network as novel determiners of *Trypanosoma cruzi* virulence. *Int J Parasitol.* 2009;39(13):1455-64.
219. Zago MP, Hosakote YM, Koo S-j, Dhiman M, Piñeyro MD, Parodi-Talice A, et al. TcI Isolates of *Trypanosoma cruzi* Exploit the Antioxidant Network for Enhanced Intracellular Survival in Macrophages and Virulence in Mice. *Infection and Immunity.* 2016;84(6):1842-56.
220. Giese V, Dallagiovanna B, Marchini FK, Pavoni DP, Krieger MA, Goldenberg S. *Trypanosoma cruzi*: a stage-specific calpain-like protein is induced after various kinds of stress. *Mem Inst Oswaldo Cruz.* 2008;103(6):598-601.
221. Osorio L, Rios I, Gutierrez B, Gonzalez J. Virulence factors of *Trypanosoma cruzi*: who is who? *Microbes Infect.* 2012;14(15):1390-402.
222. Perez Brandan C, Padilla AM, Xu D, Tarleton RL, Basombrio MA. Knockout of the *dhfr-ts* gene in *Trypanosoma cruzi* generates attenuated parasites able to confer protection against a virulent challenge. *PLoS Negl Trop Dis.* 2011;5(12):e1418.
223. Avila CC, Almeida FG, Palmisano G. Direct identification of trypanosomatids by matrix-assisted laser desorption ionization-time of flight mass spectrometry (DIT MALDI-TOF MS). *J Mass Spectrom.* 2016;51(8):549-57.

## APÊNDICES

## APÊNDICE A

International Journal of Mass Spectrometry  
418 (2017) 51–66



Contents lists available at ScienceDirect

International Journal of Mass Spectrometry



## Novel DNA coding regions and protein arginylation reveal unexplored



### *T. cruzi* proteome and PTMs

Gilberto Santos de Oliveira<sup>a</sup>, Rebeca Kawahara<sup>a</sup>, Livia Rosa-Fernandes<sup>b</sup>, Carla C. Avila<sup>c</sup>,  
Martin R. Larsen<sup>b</sup>, João Marcelo Pereira Alves<sup>c</sup>, Giuseppe Palmisano<sup>a,\*</sup>

<sup>a</sup> GlycoProteomics Laboratory, Department of Parasitology, University of Sao Paulo, Brazil

<sup>b</sup> Department of Biochemistry and Molecular Biology, University of Southern Denmark, Denmark

<sup>c</sup> Department of Parasitology, University of Sao Paulo, Brazil

#### article info

##### Article history:

Received 17 August 2016

Received in revised form 22

November 2016

Accepted 22 November 2016

Available online 9 December 2016

##### Keywords:

Proteomics

PTMs

*Trypanosoma cruzi*

MS/MS spectra interpretation Protein

arginylation

#### abstract

Chagas disease, also known as American trypanosomiasis, is a neglected tropical disease caused by the *Trypanosoma cruzi* parasite. In order to develop diagnostic and therapeutic solutions, there has been an intense investigation on the parasite biology using omics technologies such as genomics, transcriptomics lipidomics and proteomics. In particular, large scale mass spectrometry-based proteomics studies have allowed the identification and quantification of proteins and selected PTMs in different biological conditions. In this study, we investigated the unassigned MS/MS spectra commonly observed in large scale bottom up proteomics experiments looking at the *T. cruzi* (*Sylvio X10/1*) proteome. A deep proteomics data analysis using proteogenomic and unrestrictive PTMs search approaches allowed us to annotate 30% more MS/MS spectra and identify novel DNA coding regions and uncharacterized PTMs in Trypanosomatids, such as protein arginylation. Overall, this study shows: (1) the importance of assigning protein modifications, analytical artefacts and PTMs, in large-scale mass spectrometry-based proteomics data to deeply profile the trypanosomatids proteome. (2) The need of a better characterization of the influence of sample preparation steps on the identification of proteins and protein modifications. (3) The identification of novel DNA coding regions in *T. cruzi*. (4) The discovery of protein arginylation in trypanosomatids.

© 2016 Elsevier B.V. All rights reserved.

## Introduction

Chagas Disease affects millions of people around the world, mainly in Latin America. The etiological agent is the trypanosomatid flagellate parasite *Trypanosoma cruzi* which is transmitted by the triatomine insect vector. Transmission can also occur through blood transfusion, organ transplantation and congenital [1,2] which makes this disease a public health problem worldwide.

*Trypanosoma cruzi* infection is characterized by two phases, an acute phase which can be asymptomatic, and a chronic phase that affects 60–70% of patients, which is characterized by cardiac and digestive injuries. Currently, no vaccine is available on the market and treatment of Chagas' disease is prevalently based on two drugs, benznidazole and nifurtimox both with side effects [3].

*T. cruzi* strains are divided phylogenetically into six discrete typing units (DTUs) [4] named TcI–TcVI. Recently, a new strain infecting bats was characterized and named TcBat [5,6]. The genetic differences between DTUs are reflected on the geographical distribution, clinical effects and drug resistance [7,8]. TcI is found in northern Brazil and central South America and it is prevalent in the sylvatic cycle. The most characteristic manifestations of Chagas disease caused by TcI are acute and cardiac Chagas' disease, with rare cases of “mega” syndromes [9–11].

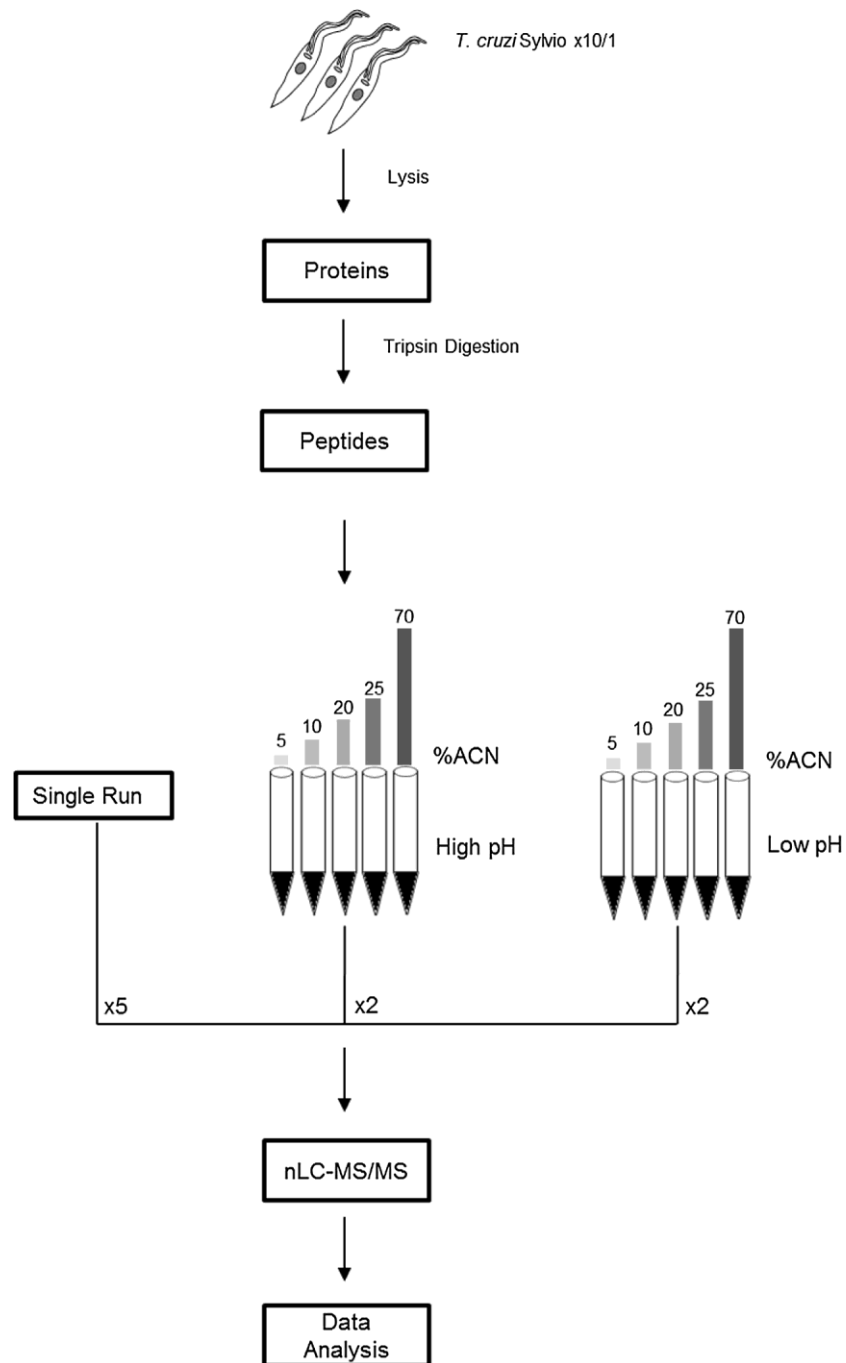
The first *T. cruzi* genome sequenced was the hybrid strain CL Brener (belonging to TcVI) in 2005, showing that the genome size is 89 Mb, containing 23.696 genes, that encode for 19.607 proteins [12]. The first assembly of the strain *Sylvio X10/1* genome was performed in 2011 [13]. Currently, the genome of *Sylvio X10/1* v2 contains 38 Mb and 41 chromosomes, corresponding to 10.861 genes of which 10.847 are protein-coding and 47 are genomic DNA. Furthermore, other strains were also sequenced, like Esmeraldo (genBank assembly accession: GCA 000327425.1), JR cl. 4 (genBank assembly accession: GCA 000331405.1), Tula cl2 (genBank assembly accession GCA 000365225.1), Marinkellei [14] and DM28c [15]

---

\* Corresponding author at: GlycoProteomics Laboratory, Department of Parasitology, ICB, University of São Paulo, Rua Avenida Lineu Prestes 1374, São Paulo, SP 05508-900, Brazil.

E-mail address: [palmisano.gp@usp.br](mailto:palmisano.gp@usp.br) (G. Palmisano).

<http://dx.doi.org/10.1016/j.ijms.2016.11.020>  
 1387-3806/© 2016 Elsevier B.V. All rights reserved.



**Fig. 1.** Proteins were extracted from lysates of *Sylvio X10/1* grown *in vitro*, digested with trypsin and these were subjected to analysis by nLC-MS/MS in different concentrations of ACN (5%, 10%, 20%, 25%, 70%) in acid and basic pH. The five biological samples ( $\times 5$ ) were subjected to LC-MS/MS with each duplicates ( $\times 2$ ), and in single run of 20, 70 and 130 min, respectively.

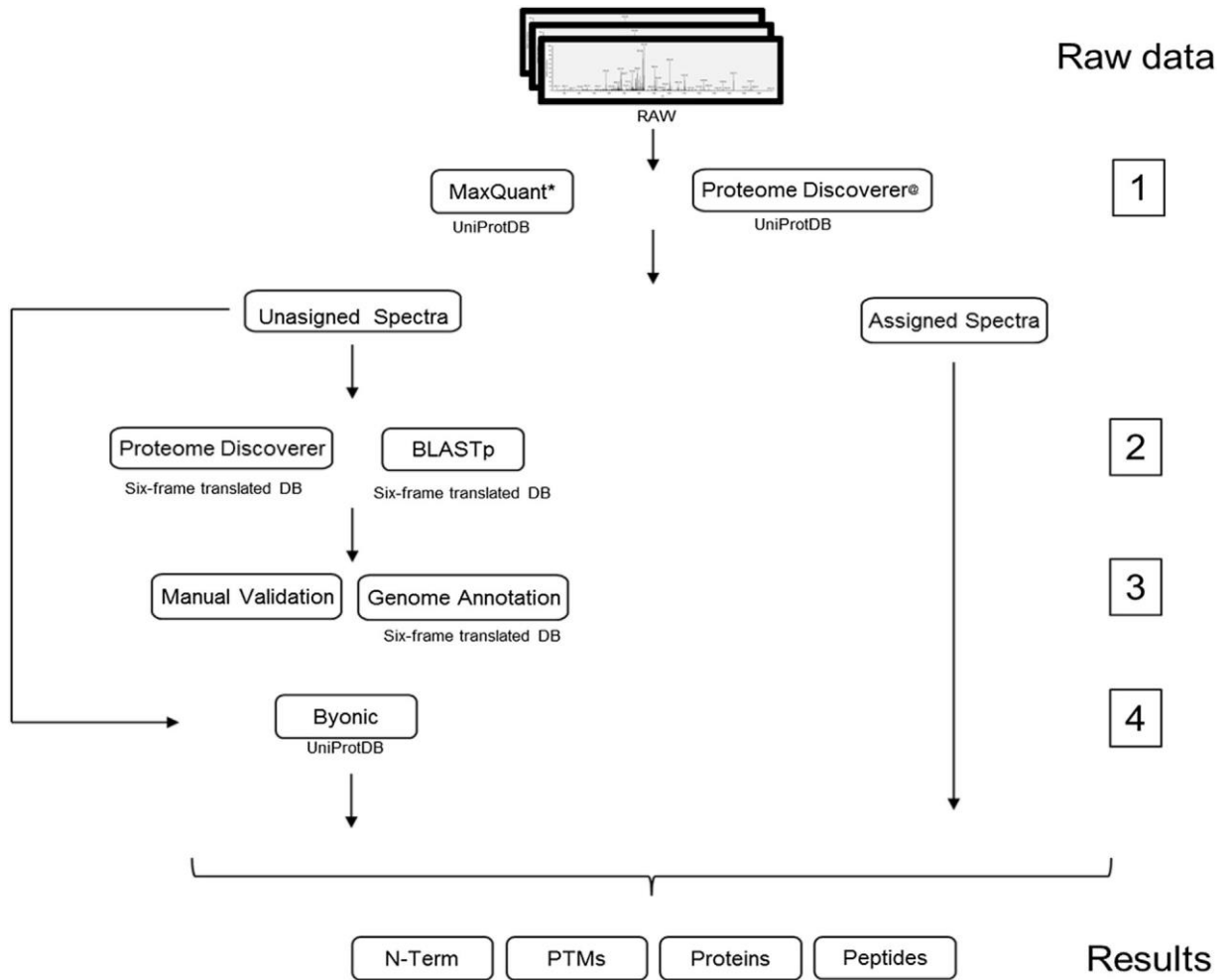
and others species, such as *Trypanosoma brucei gambiense* [16] and *Leishmania major* [17]. Beside a large effort to sequence genome of parasites causing neglected tropical diseases, the majority are still fragmented and poorly annotated.

Gene annotation can be improved with proteomics data, since it could provide information about splice variant, open read frame identification (ORFs) and patterns of gene expression at the level of protein [18–20]. In addition, proteomics could provide information about post-translational modifications (PTMs). Proteogenomic search mass spectra against genome database translated in six-frame, potentially increasing the information of protein-coding genes [21]. Additionally, proteogenomic has the

ability to identify new peptides/proteins from genome regions that primarily were not predicted to be translated, i.e., peptides in inter-genic and intragenic regions, and the identification of alternative translation start site [22,23].

Proteogenomic improved annotation confirming the ORFs expression in various organisms, for example, in eukaryotes 224 hypothetical protein found in humans [24] and confirmed new transcripts in *Drosophila melanogaster* [25]. In addition, proteogenomic was also applied to map the products of ORFs in *Mycoplasma pneumoniae* [18] and *Shewanella oneidensis* [26].

Since the first large scale proteomic analysis of *T. cruzi* [27], several mass spectrometry-based proteomics studies have been



**Fig. 2.** Deep MS data analysis. First step, *Sylyvio X10/1* MS/MS spectra were used to perform a database search in MaxQuant with specific trypsin and semi-specific, and in the Proteome Discoverer with specific trypsin, both with the UniProt database. Second step, the unassigned MS/MS spectra were extracted and filtered file was created with only the unassigned MS/MS spectra. These spectra were used to perform another search in Proteome Discoverer, but with the *Sylyvio X10/1* six-frame translated database. The peptides identified in the second step were searched using BLASTp algorithm against the six-frame translated database. Third step, the peptides identified in BLASTp were manually validated using several criteria described in the material and methods section and the gene annotation was performed. Fourth step, the unassigned spectra from step 2 were searched using the Byonic and Novor software. Finally, all peptides and proteins identified in each step were assembled.

performed using different model strains and stages of development of the parasite [28–36]. These studies demonstrated the importance of proteomics to understand the metabolic and cell signalling rewiring upon biotic and abiotic stress. These studies have helped to prioritize molecular targets for the development of therapeutic drugs. Moreover, several *T. cruzi* PTMs have been mapped [35,37–40] using well-established enrichment methods [41,42]. However, there are several PTMs that are usually undetected due to the lack of specific enrichment methods and dedicated data analysis workflows. The development of unbiased PTMs assignment has been targeted using mass tags [43], error tolerant search [44], *de novo* sequencing [45] or a combination of them [46]. Recently, it was reported a novel method based on database search with a large precursor ion mass window [47].

In this study, we analysed the proteome of *T. cruzi* Sylvio X10/1 strain in the epimastigote stage. A bottom-up proteomic workflow was applied using high-resolution mass spectrometry-based analyses for peptide sequencing. An innovative and deep data analysis workflow using a proteogenomic approach of unmatched mass spectra allowed us to identify five novel DNA coding regions and one N-terminal gene extension. Unrestrictive search significantly improved the number of matched MS/MS spectra (with a total number of 64,320 high confidence matched MS/MS spectra, of which 20,520 were matched with unrestricted modification

from 200 to +400  $m/z$ ) and allowed us to map novel protein modifications previously unknown in *Trypanosoma cruzi* such as protein arginylation with site-specific peptide mapping. The potential *T. cruzi* aminoacyltransferase that could be involved in protein arginylation was identified by sequence similarity with the ATE-like aminoacyl-tRNA-protein transferase described in *Plasmodium Falciparum*. Moreover, chemical and enzymatic artefacts due to sample preparation were identified and discussed.

Overall, this study shows: (1) the importance of assigning protein modifications, and analytical artefacts in large-scale mass spectrometry-based proteomics data (2) The need of better elucidate the influence of sample preparation steps on the identification of proteins and protein modifications. (3) The identification of novel DNA coding regions in *T. cruzi*. (4) The discovery of protein arginylation in trypanosomatids.

## Material and methods

### 1.1. *Trypanosoma cruzi* cell culture

Epimastigotes form of the *T. cruzi* Sylvio X10/1 were grown in Liver Infusion Tryptose (LIT) supplemented with 10% fetal bovine serum at pH 7.2 at 28 °C as described before [48].



### 1.2. Protein extraction and digestion

Epimastigote forms ( $1 \times 10^7$ ) were washed three times in phosphate buffered saline (PBS), pH 7.2 (8000 g for 10 min at room temperature), re-suspended in 200  $\mu$ L of lysis buffer (7 M urea, 2 M thiourea, 1 mM DTT, 50 mM Ammonium bicarbonate and protease inhibitors (Merck)) and incubated under stirring for 30 min to solubilize the proteins. Proteins were quantified using Qubit fluorometric detection (Thermo Fisher).

Proteins were reduced with 10 mM DTT (DL-Dithiothreitol – SIGMA) at 30 °C for 45 min, alkylated with 40 mM iodoacetamide (GE-Healthcare) for 30 min in the dark and diluted 10 times before

being digested with trypsin (proteomics grade Promega) in the ratio 1:50 ( $\mu$ g trypsin/ $\mu$ g protein) in 50 mM ammonium bicarbonate solution ammonium at 37 °C for 16 h. The reaction was

stopped with 1% formic acid (pH less than 3) and then the sample was desalted with C18 microcolumns using Poros Oligo R3 resin (Applied Biosystem). Peptide fractionation was performed at high pH using ammonia water and low pH using TFA, as described below.

### 1.3. Peptide fractionation

Peptide fractionation was performed using OligoR3 reversed-phase resin (Applied Biosystems) packed in a p200 pipette tip blocked with a C18 disk (Millipore). Peptides were eluted in a stepwise gradient of increasing acetonitrile concentrations. High pH fractionation was done using 0.1% ammonia (pH = 10) through the gradient described below while low pH fractionation was performed using 0.1%

TFA (pH = 3) through the gradient described below. A stepwise fractionation was implemented using different concentrations of acetonitrile (5, 10, 20, 25, 70%).

### 1.4. Mass spectrometry analysis

Peptides were loaded onto a pre-column (Reprosil-Pur C18-AQ (5  $\mu$ ; Dr. Maisch GmbH)) using an EASY nanoLC system. The peptides were eluted onto an analytical column (20 cm Reprosil-Pur C18-AQ (3  $\mu$ ; Dr. Maisch GmbH)) using the Easy-nLC HPLC. The HPLC gradient was 0–34% B solvent (A = 0.1% formic acid; B = 90% ACN, 0.1% formic acid) at a flow of 250 nL/min. Single runs were performed using 20, 70 and 120 min gradient while high pH peptide fractions were separated using a 70 min gradient. The mass spectrometer analysis was performed using the LTQ-Orbitrap Velos (Thermo Scientific). A MS scan of the mass area 400–1500  $m/z$  at a resolution of 30,000 at 400  $m/z$  for an AGC target of  $1 \times 10^6$  ions was performed. For subsequent MS/MS analysis CID was performed on the 20 most intense ions from the MS analysis. The parameters for data acquisition were: activation time = 15 ms, normalized energy = 35, Q-activation = 0.25, exclusion = available with repeat count 1, exclusion duration = 30 s and intensity threshold = 30,000, AGC target ions =  $2 \times 10^4$ . Five biological replicates were acquired in a single run while biological duplicates were fractionated as reported in Fig. 1.

### 1.5. Data analysis

Deep data analysis was performed according to Fig. 2. The data analysis workflow was divided in 4 steps: (1) raw data searched against the UniprotDB database; (2) unassigned spectra searched against the six-frame translated *T. cruzi* Sylvio X10/1 genome;

(3) matched spectra from step 2 were manually validated and matched in the ORF regions; (4) unassigned spectra from step 1 were searched using *de novo* sequencing and unrestrictive search. These steps were performed sequentially in order to improve peptide IDs and to.

#### 2.5.1. Database search against UniprotDB

Raw data were analyzed using two software, Proteome Discoverer v2.1 (Thermo Scientific) and MaxQuant [49]. Proteome Discoverer searches were performed using SequestHT and Amanda [50,51] database search engines while MaxQuant uses the embedded Andromeda search engine [52]. The database used in the first step was the *T. cruzi* Uniprot protein Database (July 2016; 54,387 entries) with the addition of common contaminants. The following parameters were used: precursor mass tolerance of 20 ppm; mass tolerance MS/MS 0.5 Da. Trypsin was selected as enzyme cleavage and carbamidomethyl cysteine as fixed modification. The variable modifications were methionine oxidation, protein N-terminal acetylation and asparagine and glutamine deamidation. Shared peptide sequences were grouped as grouped accessions proteins. Proteome Discoverer False Discovery Rates (FDR) at peptide spectrum matching (PSMs) level was calculated using the Percolator algorithm with q value equal or less than 0.01. Protein FDR was also kept at less than 1%.

Database search was also performed using MaxQuant software v1.5.3.30 [53] using Uniprot database of *T. cruzi* proteins database release (July 2016; 54,387 entries) with the addition of common contaminants. A mass tolerance of 4.5 ppm for MS and

0.5 Da for MS/MS. Enzyme specificity was set to trypsin with a maximum of two missed cleavages. Carbamidomethylation of cysteine (57.02 Da) as fixed modification, and oxidation of methionine (15.99 Da), deamidation NQ (+0.984 Da) and protein N-terminal acetylation (42.01 Da) were selected as variable modifications. Identifications with less than 1% FDR at

PSMs and protein level were accepted for further analyses. The iBAQ parameter was used as a surrogate of protein expression [54].

In order to identify processed proteins, a database search was performed using the same parameters and semi-trypsin as enzyme to account for semi-tryptic peptides. This database search was defined “closed”.

Spectra that did not produce any PSM were separated in an mgf data file for further analyses.

#### 2.5.2. Database search using six-frame translated genome

The unmatched spectra from step 1 were searched against a database consisting of the genomic sequences of *T. cruzi Sylvio X10/1* [12] after 6-frame translation by version 6.6.0 of EMBOSS’ transeq [55] of the complete contig sequences into the corresponding peptide sequences; each contig’s translation was then split based on stop signals and only peptides longer than 20 residues were kept for subsequence analyses.

Proteome Discoverer software with SequestHT and Amanda database search algorithm was used to perform database search as reported in step 1.

#### 2.5.3. Manual validation and genome annotation

Peptide spectrum matches that passed step 2 were manually validated. The criteria for MS/MS spectra filtering were the follows:

- (a) the MS scan should have less than 10% “signal contamination”;
- (b) all the most abundant ions should be assigned. (c) three consecutive y and b ions should be assigned and (d) neutral losses are not considered when are the only ions used to build the peptide sequence. Peptides that passed manual validation were mapped into the *T. cruzi Sylvio X10/1* genome using the BLAST algorithm as described below. All BLAST searches had low complexity filter and

composition-based statistics options turned off to assure that the whole sequences would match their targets and results were manually analyzed to eliminate false positive matches.

Peptide sequences identified in the database searches against the 6-frame translated *T. cruzi Sylvio X10/1* genome and filtered by manual validation were analyzed, together with available annotation data for this genome, with TBLASTN [56] (demanding 100%

sequence identity and 100% query length coverage) and bedtools

[57] in order to identify any peptides that partially overlap an already known gene. Also, given the fragmentary nature of the *T. cruzi Sylvio X10/1* genomic sequences used, contigs where the peptides mapped were searched against NCBI's NR protein database using BLASTX, for the identification of peptides matching genes that have not been predicted in this genome due to sequence incompleteness.

The open reading frames in *T. cruzi Sylvio X10/1* corresponding to each of the final "intergenic" peptides were identified and searched by BLASTP against NR to identify any previously predicted trypanosomatid proteins. Next, these same ORFs were searched by TBLASTX against all genomic sequences available for trypanosomatids (as well as *Bodo saltans* and *Parabodo caudatus*), in order to find any possible gene candidates for further analyses. Regions matching the ORFs (plus 2000 bases on either side of the match) were then isolated with the aid of custom Perl scripts and EMBOSS' extractseq, translated as above, and analyzed by EMBOSS' getorf ORF prediction and BLASTP searches to identify the genomic fragments corresponding to the peptides of interest.

All sequences putatively corresponding to each novel *T. cruzi Sylvio X10/1* gene were aligned using Muscle v. 3.8.31 [58] and manually examined and edited in Seaview v. 4.5.4 [59] to remove badly aligned and incomplete sequences. The putative translation start of each novel gene was determined by manual analysis of conservation along the corresponding translated sequence alignments.

#### 2.5.4. Unrestricted PTMs search analysis

Unknown or unexpected modifications were determined by searching the non-identified filtered spectra (.mgf) against the Uniprot *Trypanosoma cruzi* fasta database as in step 1 using Byonic software v.2.6.46 (<http://www.proteinmetrics.com/> Protein Metrics Inc.). Searches were performed with the following fixed modifications: precursor mass tolerance of 10 ppm, product ion mass tolerance of 0.5 Da, carbamidomethylation Cys, and fully trypsin specific cleavage with a maximum of two missed cleavages. Searches were also conducted with oxidation of methionine (15.994 Da) as variable modification. Initial searches allowed a wild card modification (any mass delta from 200 to +200 Da on any one amino acid residues). Subsequently, a Byonic search was performed using a wild card modification from +200 to +400 delta mass.

High confident PSMs were selected based on the 2D FDR system developed in the Byonic software [60]. Specific protein modifications were further characterized considering delta mass shift with less than 10 ppm accuracy and MS/MS fragmentation spectral quality as reported in step 3. This database search is defined as "open".

#### 2.6. Bioinformatics analysis

The five newly annotated coding DNA sequences were further investigated for possible conserved domains that could elucidate their function. The Conserved Domain Database [61] (CDD) and Motif Scan [62] was used to perform the searches. Moreover, the presence of signal peptides was evaluated using SignalP v4.1 Server [63], which consists in predicting the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. Proteins that contained a signal peptide according to

SignalP prediction were further evaluated using SecretomeP [64].

Prediction of possible PTMs was conducted using the software-based family Group Prediction System (GPS). For phosphorylation we used Group-based Phosphorylation Scoring v3.0 (GPS) [65]; For acetylation was used Prediction of Acetylation on Internal Lysines [66] (PAIL); GPS-Small Ubiquitin-like Modifier v1.0

[67] (GPS-SUMO) for Sumoylation; GPS-Pupylation [68] (GPS-PUP) for pupylation; Tyrosine Sulfation Protein [69] (GPS-TSP) for tyrosine sulfation; GPS-SNO [70] for S-nitrosylation. For lipid

modifications was used GPS-Lipid [71]. In addition, was used the softwares available in the Center for Biological Sequence analysis (CBS), such as (NetNGlyc 1.0 Server) (<http://www.cbs.dtu.dk/services/NetNGlyc/>) for N-glycosylation; NetOGlyc 4.0 [72] Server for O-glycosylation (<http://www.cbs.dtu.dk/services/NetOGlyc/>); YinOYang 1.2 [72] for O-GlcNAc (<http://www.cbs.dtu.dk/services/YinOYang/>). It should be noted that the entire ORF was imputed for bioinformatics prediction.

#### 2.7. Phylogenetic analysis of the putative arginyltransferase

The *Plasmodium falciparum* ATE-like aminoacyl-tRNA-protein transferase (ATEL1) protein sequence (accession number Q195H6) was used as a query to identify putative candidates for this function in kinetoplastids and other organisms. Identification was performed by BLASTP search (E-value threshold of 1E-3) against the complete NCBI non-redundant database. All matches were extracted and aligned with Muscle v.3.8.31, and alignments were manually examined with Seaview v.4.5.4 in order to identify and remove badly aligned sequences. The remaining sequences were analyzed by RAxML v.8.2.9 [73] in order to infer the maximum likelihood phylogeny. The substitution model was chosen automatically by the program, using gamma-distributed rate heterogeneity rates, and an automatically determined number of bootstrap pseudoreplicates were run to assess statistical support for the individual branches. Resulting trees were processed and drawn in TreeGraph2

[74] and Dendroscope [75].

1.6. Understanding the mass spectrometric behavior of the *T. cruzi* (Sylvio X10/1 strain) proteome

*Trypanosoma cruzi* cells from Sylvio X10/1 strain in epimastigote stage were lysed and proteins were extracted, digested with trypsin and analyzed using nanoflow LC–MS/MS. In order to deeply sequence the proteome, tryptic peptides were fractionated using reversed-phase chromatography in acidic (pH = 3) and basic (pH = 10) conditions with a stepwise acetonitrile concentrations (5%, 10%, 20%, 25% and 70%). Due to that, we implemented an offline RP–RP 2D chromatographic system using two mobile phase additives such as TFA for low pH and ammonia for high pH in the first dimension, separately. The intrinsic complexity and dynamic range of the *T. cruzi* proteome required peptide fractionation using an orthogonal chromatographic system and allowed us to reduce peptide co-elution and ion suppression and to improve protein IDs [76,77]. Fractionated samples were prepared in a biological duplicate, while five biological replicates (x5) were directly injected in the nLC–MS/MS system without fractionation, as shown in Fig. 1.

The raw mass spectrometric data were analyzed using the data analysis workflow illustrated in Fig. 2. Initially, a database search approach using the *T. cruzi* UniProt database was performed using MaxQuant and Proteome Discoverer software. For MaxQuant search trypsin with specific and semi-specific cleavage was used. A summary of the identifications is reported in Table 1.

Table 1 shows the identifications obtained from each software; Proteome

Discoverer identified 169585 MS/MS spectra out of 493492 whereas MaxQuant identified 124661 and 124732 MS/MS spectra out of 535437 using trypsin and semi-trypsin specificity, respectively (Table 1). It is worth mentioning that there is redundancy for features identified by both programs. Thus, of the total analysis, 17752 peptides are unique in the identified by Proteome Discoverer, Supplementary Table S1a, while 17999 and 18030 are unique in MaxQuant trypsin and semi-trypsin specificity, respectively, with 1% FDR, Supplementary Table S1b. We noted that there

**Table 1**Deep data analysis of the *T. cruzi* (Sylvio X10/1 strain) proteome.

Data	Database	Software	Features	Unique peptides/Proteins	Observations
<b>1</b>					
RAW	UniProt	Proteome Discoverer	169586/493492 <sup>a</sup>	17752/3851	Trypsin
RAW	UniProt	MaxQuant	124661/535437 <sup>a</sup>	17758/3647	Trypsin
RAW	UniProt	MaxQuant	124732/535437 <sup>a</sup>	17776/3636	Semi-trypsin
<b>2</b>					
Filtered	Six-frame translation	Proteome Discoverer	3886/361647 <sup>b</sup>	354	
<b>3</b>					
Peptide sequences after BLASTp		Manual Validation	239/354 <sup>c</sup>	239	
Peptide sequences validated	Six-frame translation	Genome annotation	5/239 <sup>e</sup>	5	
<b>4</b>					
Filtered	UniProt	Byonic <sup>d</sup>	64320/361647	8327/2771	
Total <sup>e</sup>				23374/5262	

©Peptides.

#High confident sequences.

<sup>a</sup> Redundant features.<sup>b</sup> PSMs.<sup>c</sup> Manual validation used stringent criteria such as: (1) all high intense fragments should be annotated, (2) three subsequent ions y and b.<sup>d</sup> Only modified peptides.<sup>e</sup> Total unique peptide sequences without modifications. The wild card search in Byonic is excluded.

was no significant difference in the unique peptides identified from the trypsin and semi-trypsin searches in the MaxQuant, indicating very low nonspecific cleavage due to sample degradation or chymotryptic activity in the trypsin batch.

Using the results from the MaxQuant database search, we observed that the most expressed proteins are related to glycolysis 21%, heat shock protein (HSP) 9%, histones 9% and tubulin 6% and other highly expressed proteins are involved in carbon, arginine, proline metabolism and biosynthesis of secondary metabolites, Supplementary Fig. S1A. In the low expression range, proteins are mostly unannotated (68%) or uncharacterized (32%), Supplementary Fig. S1B. The high number of unannotated proteins in low abundance can contribute to the lack of biochemical information and calls for more studies. The identified proteins can be visualized in the Supplementary Table S1c.

In total, the number of matched tandem mass spectra, using database searching against a reference database such as the UniprotKB, was 34%. This is slightly higher compared to the first large scale proteome analysis of *T. cruzi* that was reported to be 12% [29]. In order to increase the identification rate of the unassigned MS/MS spectra, we performed a deep data analysis strategy as reported in Fig. 2. In particular, MS/MS spectra that were not identified using a database search approach with the UniProtKB database were filtered, extracted and converted to mgf format. This filtered file was then used to perform a database search against six-frame translated

*T. cruzi* *Sylvio X10/I* database, step 2 and 3. Moreover, the filtered unassigned spectra were searched with a wide mass tolerance

using the Byonic software against the UniProt database.

The database search using the six-frame translated *T. cruzi* genome allowed us to identify 3886 of 361647 MS/MS spectra. These 3886 PSMs matched to 354 unique peptide sequences Supplementary Table S2a. The MS/MS spectra of these 354 unique peptides were subjected to manual validation using the rules reported in the Materials and Methods section [78]. A total of 239 unique peptide sequences passed the manual validation filter, Supplementary Table S2b and Table 1.

Moreover, we performed similarity searches of the 239 manually curated peptides that did not match predicted *T. cruzi* *Sylvio X10/I* against the six-frame translation of this organism's contigs, and we found that all peptides matched, completely and perfectly, at least one genomic region. However, given the very fragmented nature of the available genome sequence, which leads to many



unpredicted genes at the contigs' ends, another filtering step was necessary. In this case, all contigs were searched by BLASTX against the complete NCBI non-redundant (nr) protein database, in order to find any fragments of genes that were missed in the annotation most likely due to genomic sequence incompleteness. Next, bedtools was used to find all intersections between the coordinates of nr matches (i.e. unpredicted genes) and peptide matches on the contigs. All but five of the 239 manually curated peptides were found to be part of an incompletely sequenced and annotated gene. The five peptides were: pep76, pep198, pep206, pep304 and pep314, Supplementary Table S2b. Open reading frames of *T. cruzi* Sylvio X10/1 genome containing these novel intergenic peptides were identified and are reported in Supplementary Table S3a and b.

#### 1.7. Novel protein coding regions in the *T. cruzi* Sylvio X10/1 strain genome

In order to understand the nature of these novel DNA coding regions, we applied several bioinformatics tools to calculate their molecular weight and predict their subcellular localization and PTMs, Table 2. These novel ORFs codes for relatively small proteins with a theoretical molecular weight between 6 and 23 kDa. None of them contains a conserved domain that could indicate their biochemical function. The predicted subcellular localization was diverse from nucleus (pep76), cytoplasm (pep198 and pep304), endoplasmic reticulum (pep206) and Golgi (pep314). Interestingly, pep314 and pep198 were predicted to be secreted with a signal peptide. The

number of transmembrane domains was low in the majority of the novel proteins indicating their soluble or membrane-associated nature. Furthermore, we looked at predicted PTMs, such as phosphorylation, N- and O-glycosylation, O-GlcNAcylation, lysine acetylation, SUMOylation, tyrosine nitration, and lipidation. Phosphorylation sites were predicted in all the 5 proteins with pep198, pep206, pep304 and pep314 containing predicted tyrosine phosphorylation sites. Tyrosine phosphorylation is an important regulator of signal transduction with the ability to switch on/off specific signaling cascades [79] and protein kinase inhibitors have been suggested with the potential for treating parasitic diseases [80,81]. Interestingly, several predicted phosphorylation sites were also predicted to contain an O-GlcNAc modification. None of the novel genes were predicted to contain N-

**Table 2**  
Predicted Post-translational Modifications of novel DNA coding regions in *T. cruzi*.

Co/Post-translation modifications	Novel DNA coding regions				
	pep76 ADWP 02021363 139	pep198 ADWP 02017124 21	pep206 ADWP 02006304 64	pep304 ADWP 02015018 214	pep314 ADWP 02013302 220
Predicted MW <sup>a</sup>	6596.41	23116.50	7820.11	9217.21	9680.47
SecretomeP <sup>b</sup>	0.944	0.641	x	0.885	0.883
Subcellular <sup>c</sup> localization	Nucleus	Signal peptide Cytoplasm	Endoplasmatic reticulum; mitochondrion	Cytoplasm	Signal peptide Golgi; Mitochondrion
Transmembrane domains <sup>d</sup>	0	3	1	0	0
Phosphorylation	S-4, S-15, T-17, T-32, S-36, T-46, S-55, T-57	S-5, T-7, T-12, S-21, T-25, S-27, T-55, S-56, T-96, T-102, T-104, S-122, Y-132, Y-175, T-185, T-193, T-196	S-9, T-11, S-26, S-31, Y-39, Y-42, S-43, T-44, S-55	T-3, T-7, Y-16, S-19, S-29, Y-33, Y-34, Y-44, S-58, S-61, S-66, T-69, S-76, S-78, S-79	S-4, S-29, T-46, Y-55, S-57, T-61, S-63, S-69, T-73, S-77
N-Glycosylation	x	x	x	x	X
O-Glycosylation	x	x	x	x	X
O-GlcNAc	S 4; T 32; T 57	193 T; 196 T	S 31	S 29; S 76	T 73
Acetylation	K 6, 9, 25, 29, 30, 31, 38	K 9, 47, 89, 106, 108, 200, 201	K 68	K 5, 6, 8, 9, 10, 12, 14, 17, 47	K 59, 71, 72, 75, 76
Methylation	x	x	R 70	K 12	X
SUMOylation	50-54 NIFTEEQLIALL SDTH***	142-146 DVKLPEA VGILGPNGEVIP	59 K	x	72 K
Pupylation	x	x	x	K 43	X
Tyrosine	x	x	x	Y 16	X
Nitration					
S-nitrosylation	x	C 2	x	C 28	C 1, 33
Palmitoylation	x	C 2, 4, 16, 168	x	x	C 1
Farnesylation	x	x	x	x	X
Geranylgeranylation	x	x	x	x	X
Myristoylation	G 8	x	x	x	X

£ Prediction of methylated arginine and lysines was performed using the Methylation Modification Prediction Server (<http://www.bioinfo.tsinghua.edu.cn/~tigerchen/memo/>).

<sup>a</sup> Molecular weight was predicted using the Compute pI/MW webtool at the expasy site ([http://web.expasy.org/cgi-bin/compute\\_pi/pi\\_tool](http://web.expasy.org/cgi-bin/compute_pi/pi_tool)).

<sup>b</sup> Non-classically secreted proteins were predicted based on the NN-score. Score exceeding the threshold of 0.5 for bacterial sequences and 0.6 for mammalian sequences were considered. The recommended thresholds are 0.5 for bacterial sequences and 0.6 for mammalian sequences.

<sup>c</sup> Amino Acids. 2007 Jul;33(1):57-67. Epub 2007 Jan 19. Euk-PLoc: an ensemble classifier for large-scale eukaryotic protein subcellular location prediction. Shen HB1, Yang

J, Chou KC.

<sup>d</sup> The number of transmembrane domains were predicted using the TMHMM webtool (<http://www.cbs.dtu.dk/services/TMHMM/>).

and O-linked indicating that these proteins are not predominantly located to the cell membrane. In a recent study we quantitatively mapped the *T. cruzi* N- and O-linked glycoproteome showing widespread changes during life stage development [40].

All five novel proteins were predicted to contain a lysine acetylation on more than one site. The importance of acetylation in *T. cruzi* was shown by using inhibitors of sirtuins, (deacetylases), that prevented parasite growth and multiplication after host-cell invasion [82]. Pep206 and pep304 were predicted to be methylated in the R70 and K12, respectively. Lysine and arginine methylation has been reported on *T. cruzi* histone H4 at position K18 and R53 [83]. Another PTM of great importance in eukaryotes is SUMOylation, which is characterized by addition of an ubiquitin-like protein that is covalently attached to lysine residue in the target protein by an isopeptide bond and has been reported to be involved in several roles such as cell cycle progression, subcellular localization, protein-protein interaction, transcription and DNA repair [84]. All five proteins analyzed had predicted SUMOylated lysines. Bayona et al. showed by proteomic and immunological methods, using polyclonal anti-TcSUMO, that SUMOylation is present in all the *T. cruzi* life stages parasite and involved in several processes such as

chromatin organization and remodelling, DNA repair and transcriptional processes [84].

Recently, it was described a prokaryotic ubiquitin-like protein, Pup (Rv2111c), which was specifically conjugated to proteasome substrates in the pathogen *Mycobacterium tuberculosis* and involved in protein degradation [85]. The pep304 had a predicted K 43 pupylation. Interestingly, in our wide tolerant search we found 28 peptides with a delta mass shift that could correspond to pupylation addition to lysine. However, careful considerations are needed as reported below.

After macrophage infection, the *T. cruzi* trypomastigote stage is exposed to several protein oxidation reactions due to the NO and its derivatives production [86]. Two of the post-translational oxidative modifications are cysteine nitrosylation and tyrosine nitration. Tyrosine nitration caused by *T. cruzi* in host proteins might promote the survival of the parasite [87]. Recently, these two PTMs were found to be modulated by *T. cruzi* incubation with extracellular matrix [37]. The novel *T. cruzi* proteins identified in this study were predicted to be nitrated in Y16 (pep304) and nitrosylated in C2 (pep198), C28 (pep304) and C1, 33 (pep314).

Subsequently, protein lipid modifications were predicted in these five novel proteins. In particular, protein palmitoylation, far-nesylation, myristoylation and geranylation were investigated. The

*T. cruzi* phosphoinositide (PI)-specific phospholipase C gene (TcPI- PLC) was shown to be expressed during the trypomastigote to amastigote differentiation and it was N-myristoylated *in vivo* [88]. Protein palmitoylation was predicted on C 2, 4, 16, 168 (pep198) and on C1 (pep314). Myristoylation was predicted on G on pep76. A sequence similarity search of these novel proteins showed high homology in Trypanosomatids, although with some differences. In particular pep76, pep198 and pep314 are conserved in Trypanosomas except *T. brucei*. Fig. 3 and Supplementary Figs. S2 and S5, respectively. On the other end, pep206 and pep304 are conserved across all Trypanosomatids including Leishmania species,

Supplementary Figs. S3 and S4.

Taken together these data show the five *T. cruzi* DNA coding regions identified in this study are translated into relatively small proteins that are conserved in all Trypanosomatids and are predicted to have several PTMs.

In the first large scale mass spectrometry-based analysis of the *T. cruzi* proteome, 12% of the MS/MS spectra were assigned (17225/139147) using database search against *T. cruzi* (CL Brener strain) reference genome [27]. In our case 169586/493492 (34%) MS/MS spectra were assigned using database search against a reference genome, Table 1-step 1. In a recent study it was reported

a 50% success in MS/MS spectra assignment using new generation Orbitrap analyzer [89]. After database search using the six-frame translated *T. cruzi* genome, we observed that 357761/493492 MS/MS spectra still remained unassigned, Table 1-step 2, 3. These unassigned spectra correspond to 65% of the total MS/MS acquired, meaning that 35% of the spectra were assigned after database search using UniprotDB and six-frame translated genome. Since using a six-frame translated genome improved the number of PSMs by 1%, we questioned if there were unknown/unassigned modifications of peptides that could significantly contribute to the number of unassigned MS/MS spectra.

#### 1.8. Modified peptides identified in 200 Da precursor ion window revealed "neglected" PTMs in *T. cruzi*

After exporting the unassigned spectra from the database search using Proteome Discoverer, we performed a wide-tolerance database search (using a 200 Da window) aiming at identifying unanticipated as well as novel modifications in the peptide sequences from the *T. cruzi* proteome database. This analysis allowed us to identify 127,120 spectra assigned to peptides in 3908 proteins with a time cost of more than 10 days using our computational facility (data not shown). From these, we filtered only for the high scoring matched peptides (PEP 2D score < 0.001), which resulted in 48,893 assigned spectra, Table 1-step 4. Besides, spectra matched to peptides in the reverse database (total 27 spectra, 0.1% FDR) or identified in the contaminant database (total 552 spectra) were excluded. The peptides identified with "Wildcard" as variable modification (200 Da) were obtained as reported in Materials and Methods section, resulting in 21,388 spectra, Supplementary Table S4, which was used for the further analysis. These modified peptides covered

a total of 1500 proteins. Interestingly, 53 proteins were able to explain 50% of all modified peptides, Supplementary Fig. S6a; among them, 21 were found in the top 100 most abundant proteins (e.g. Beta tubulin, Pyruvate phosphate dikinase 1, Paraflagellar rod protein 3, putative, Tyrosine aminotransferase, Methyltransferase, ATP synthase subunit beta, Glutamate dehydrogenase, Heat shock protein (HSP70) and Histidine ammonia-lyase). Additionally, using the unrestricted database search, 743 proteins that were not identified in the database search were identified

(25%), resulting in a total number of identified proteins of 3751 in the *T. cruzi* proteome (Supplementary Fig. S6b).

We compared the identified modification in the proteome from

*T. cruzi* with the recently published dataset that performed an “open” database search (500) in a shotgun proteomic data from human HEK293 cell line [47]. Firstly, we compared the frequency distribution of the mass change identified in 149457 spectra (only spectra with Omass modifications > 0 Da and in range of 200 Da) from Chick et al., and from our data. We observed that the distribution of the Omass modifications is very similar between Human and *Trypanosoma cruzi* dataset, especially in the most abundant modifications (Fig. 4). Slight differences were observed when the Omass range was divided into smaller bins (Fig. 4a–d). For example, the modification +152 Da was mainly observed in the *T. cruzi* compared to the human dataset, whereas +183 Da was mainly observed in the Human dataset (Fig. 4d). We observed the +152 Da modification occurred with more frequency in cysteine, but because all cysteine had previously set as fixed carbamidomethylation modification (+57 Da), the actual modification is +209 Da. We found that +209 Da is annotated in the Unimod.org database as “carbamidomethylated DTT modification of cysteine”, which was reported before to be an artefact due to excess of DTT used during sample preparation [90]. In the Human dataset, the +183 Da modification was reported to be a aminoethylbenzenesulfonylation that occurs when using AEBSF (Pefabloc) as a serine protease inhibitor [47]. Therefore, specific differences in the sample preparation step may affect the number of

peptides with unanticipated chemical modification.

The percentage of each  $O_{mass}$  modification per amino acid was

calculated and the distribution of these modifications in the twenty different amino acids was visualized as a heat map using the top 40 most frequent modifications observed. We also showed the annotation available in the Unimod.org database for each  $O_{mass}$  bin in Fig. 5.

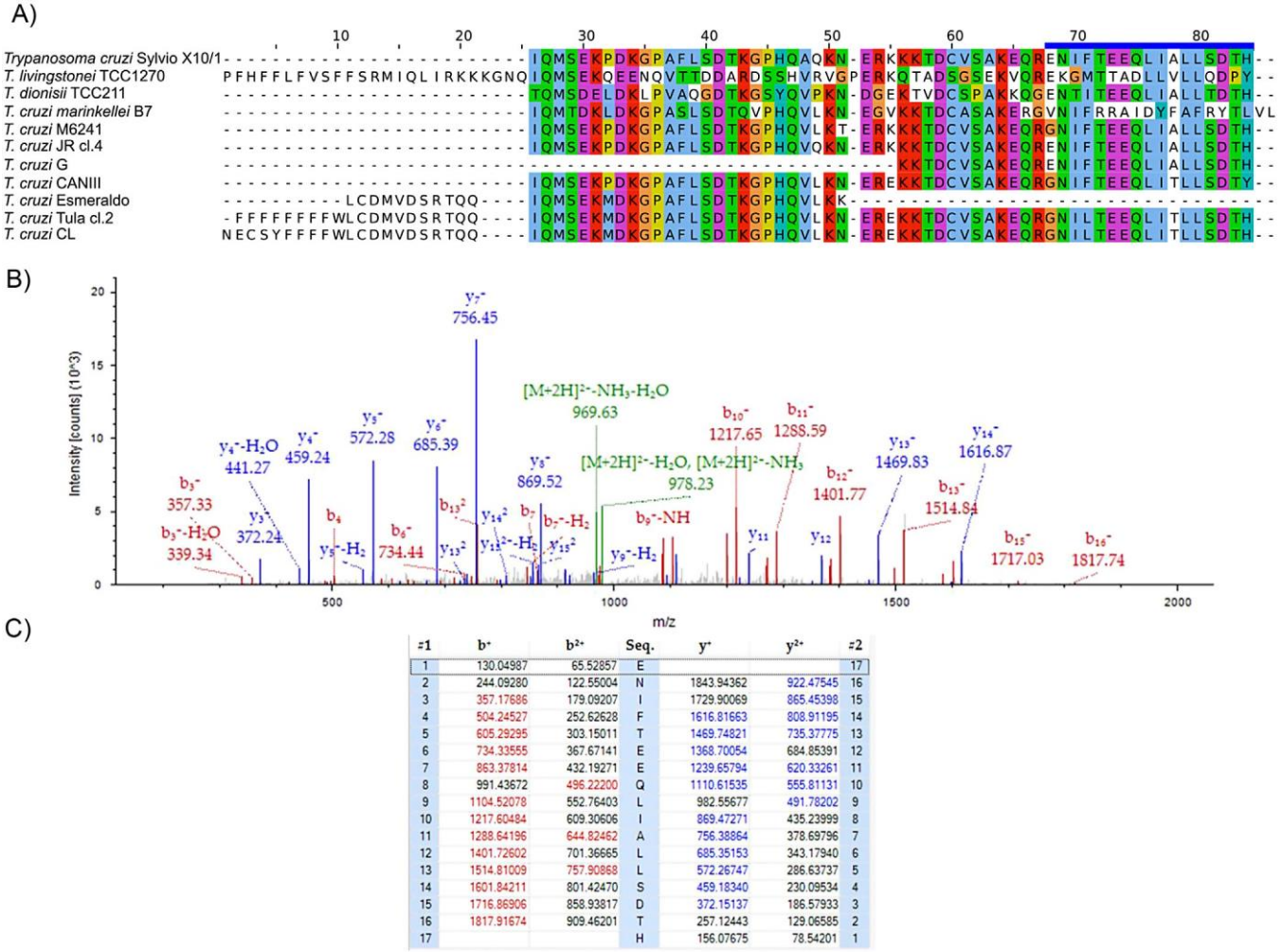
Known PTMs were detected such as phosphorylation ( $O_{mass} = +79.96$ , 355 counts, mainly in serine), acetylation ( $O_{mass} = +42.01$ , 890 counts, mainly in serine), dimethylation ( $O_{mass} = +28.03$ , 647 counts, mainly in lysine residues) etc. Amino acid substitutions were also observed, for example in Ser- > Tyr ( $O_{mass} = +76.0$ , 127 counts), Thr- > Ala ( $O_{mass} = +30.0$ , 89 counts), Ser- > Asn ( $O_{mass} = +27.0$ , 90 counts), Leu/Ile- > Val substitution ( $O_{mass} = 14.0$ , 181 counts) (Fig. 5).

Known chemical modifications were identified, for example oxidation with increased frequency on methionine and tryptophan ( $O_{mass} = +16.0$ , 1105 counts), carbamylation on lysine and methionine ( $O_{mass} = +43.0$ , 946 counts) and deamidation on asparagine ( $O_{mass} = +1.0$ , 1172 counts) (Fig. 5).

We found four abundant modifications with transpeptidation in lysine and arginine-related annotation in Unimod.org database. The  $O_{mass}$  corresponding to a predicted modification of the addition of lysine ( $O_{mass} = +128.1$ ) and arginine ( $O_{mass} = +156.1$ ) was identified in 709 and 607 spectra, respectively, whereas loss of lysine ( $O_{mass} = 128.1$ ) and arginine ( $O_{mass} = 156.1$ ) due to transpep-

tidation was identified in 72 and 354 spectra (Fig. 5).

Modification occupancy was calculated by dividing the number of spectra counts from modified peptides by the sum of the total spectra counts from the modified and respective non-modified peptide, which was identified using MaxQuant for database search. The distribution of the % values was compared using different datasets: total number of peptides identified in the modified and non-modified database search, peptides with +80 Da modification (predicted by Unimod.org database to be a phosphorylation), peptides with +43 Da modification (predicted to be a carbamylation) and peptides with +156 and +128 Da (predicted to be an addition of arginine and lysine due to transpeptidation, respectively) (Supplementary Fig. S6c). We observed that the modifications are generally in low stoichiometry compared to the non-modified peptides,



**Fig. 3.** (a) Sequence homology of the pep76 protein, (b) MS/MS spectrum and (c) identified product ions of the peptide (ENIFEEQLIALLSDTH) identified using database search against the *T. cruzi* (SylvioX10/1 strain) six-frame translated genome.

but for PTMs which are known to be low abundant, for example, phosphorylation, we found that the occupancy is increased, meaning that the modification is likely to be specific with more peptides found in the modified form compared to the non-modified form (Supplementary Fig. S6c). For a known chemical modification derived by urea (+43, carbamylation) which is frequently used in proteomics sample preparation, we found modified low stoichiometry compared to the non-modified form (Supplementary Fig. S6c). The distribution of the +156.1 and +128.1 modification was very similar and most of the peptides were identified with less than 20% of modification occupancy (Supplementary Fig. S6c). It is important to remember that this is an estimation analysis of modification occupancy because we might have different ionization properties in the non-modified and modified peptides.

Additionally, we also found 83 peptides with a +129.11 modification, which is annotated in the Unimod.org database as a potential mono-glutamyl modification (Supplementary Table S5).

Finally, we performed additional unrestricted search analysis using a delta mass window from 200 to 400 Da. From this analysis, 1482 peptides were identified with high score (PEP 2D < 0.001) (Supplementary Table S6). Among the most abundant modification in this range, we identified the delta mass of +243.1 in 40 peptides, with increased frequency in the amino acid K (28/40).

This modification is annotated in Unimod.org database as pupylation, which is known as a post-translational modification in prokary-

otes, but have never been described in *Trypanosoma cruzi*. Tryptic peptides with a pupylated lysine contain the addition Q\*GG where Q\* is a deamidated glutamine. However, careful analysis of MS/MS spectra should be carried out to discriminate peptides containing sumoylated lysine (QGG) and deamidation (Q/N). The peptides identified in this study, carrying the delta mass of +243.1, contained one or more Q/N, so it is not possible to distinguish these peptides from formerly sumoylated peptides. Further analyses are needed to confirm this observation. Moreover, a BLASTp search using the Prokaryotic ubiquitin-like protein Pup from *Mycobacterium tuberculosis* against Trypanosomatids gave no match (data not shown) indicating that these peptides might be SUMOylated (QGG) and deamidated on another residue in the same peptide. We also observed an abundant modification with delta mass of +209.01 on tryptophan, (40 out of 44 peptides) opening new avenues to unknown modifications occurring in biological systems.

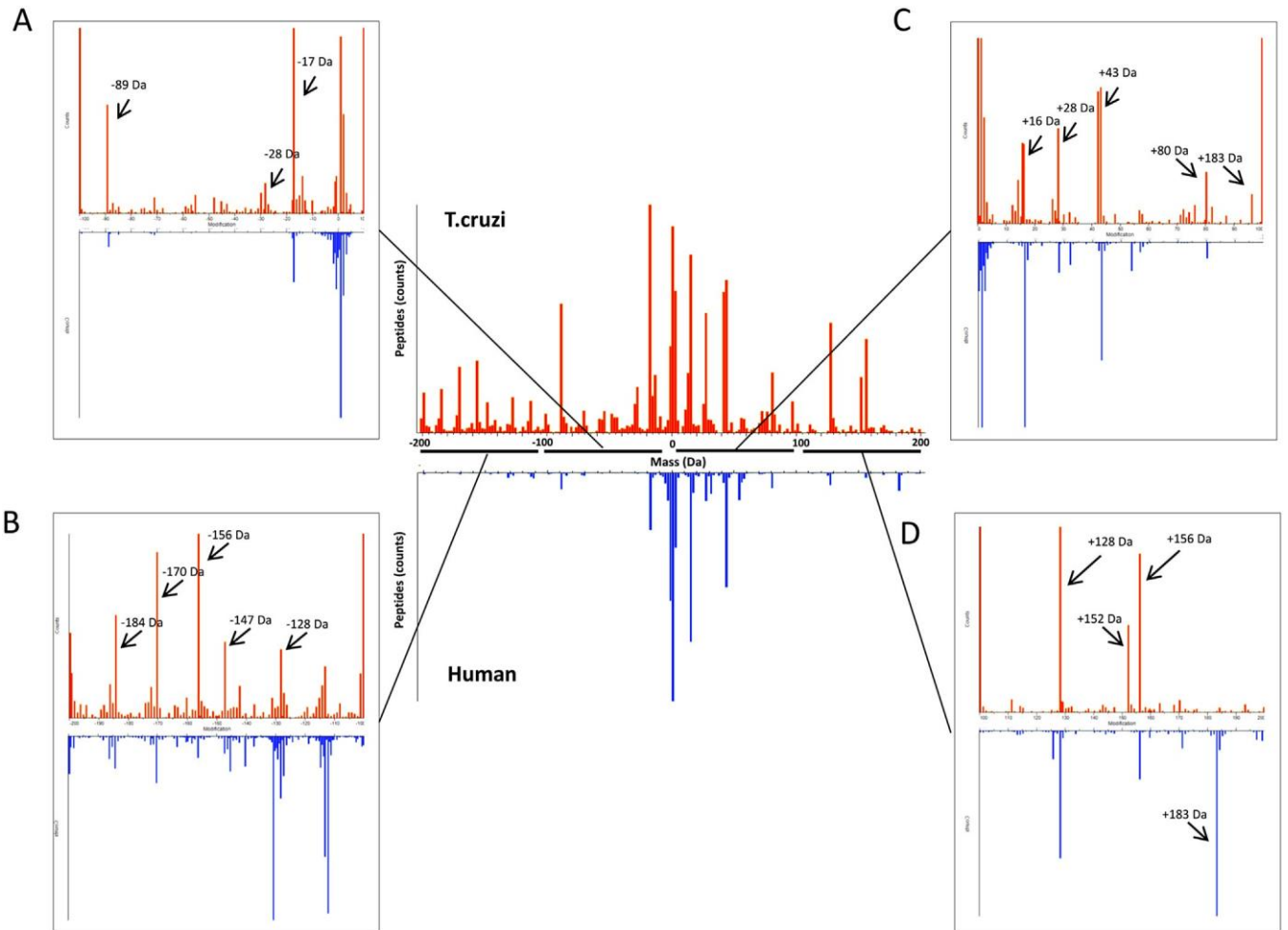
#### 1.9. Protein arginylation and transpeptidation contribute to the addition of 156.1 m/z

We were interested to investigate deeper the modification of

+156.1, which corresponds to arginine addition. Arginine addition can be due to: (1) transpeptidation or (2) protein arginylation.

In this study, a bottom up proteomic approach was applied to study the *T. cruzi* proteome. Although several proteolytic enzymes





**Fig. 4.** OMass (observed-theoretical) distribution comparison for the 21,288 modified peptides from this study and 14,947 modified peptides from Chick et al. Inset data for all peptides with a zoomed-in view in delta mass modification from 0 to -200 Da (A); -100 to -200 Da (B); 0 + 100 Da (C) and +100 to +200 Da (D). The arrows highlights the most abundant modifications found using unrestricted database search.

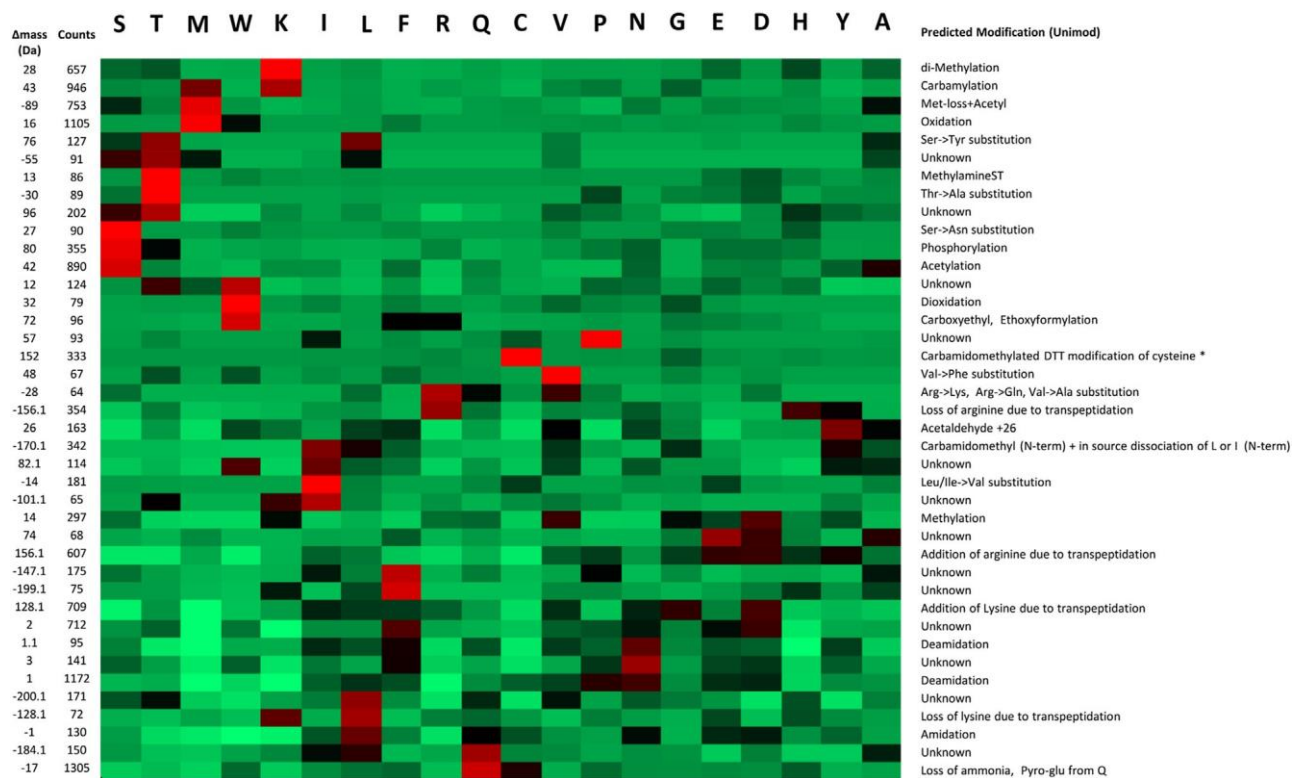
have been described to improve sequence coverage [91,92], trypsin is still the most widely used since it generates with relatively high specificity peptides with an average of 14 amino acids, based on the Uniprot human protein database, which is ideal for LC-MS/MS. In addition, tryptic peptides contain minimum two charges at the N-terminus and the C-terminal Arg/Lys, respectively. However, this enzymatic process produces side reactions such as missed cleavages [93], trypsin auto-digestion, semi-tryptic peptides due to chymotryptic activity [94,95] and transpeptidation [96]. Transpeptidation reaction results in the addition/substitution of an amino acid or an oligopeptide. In a previous study using a set of standard proteins, it was shown that transpeptidation reaction can generate the addition of a single arginine (+156.101) and lysine (+128.101) to the N-terminal of peptides. Moreover, peptides with the addition of two or more amino acids were detected [54] and the percentage of the transpeptidated peptides compared to the unmodified ones was relatively low (1–10%).

In order to discriminate transpeptidation from protein arginylation, which is a known PTM catalyzed by the ATE1 enzyme in humans, we started comparing the site specific distribution of the  $O_{mass} = +156.1$  on the different amino acids detected in our large-scale mass spectrometry-based proteomics approach on *T. cruzi* proteins. We observed that this modification occurred mainly in the amino acids E (14%) and D (10%) (Supplementary Fig. S7A and Table S7), which is in accordance with previous studies that reported protein arginylation on these amino acids [97,98]. Besides,

59% of all matched spectra contained the modification in the peptide N-terminal, 37% occurred within the peptide sequence and 4% occurred in the C-terminal (Supplementary Fig. S7B and Table S7). As described above, N-terminal addition of arginine or lysine to peptides has been observed as transpeptidation side reaction of trypsin [96]. Besides, Xu et al. specified mass ambiguities which can mimic the addition of arginine onto the N-terminus of peptides and can confuse the results of the mass spectrometry analysis. For instance, carbamylation of N-terminal Leu and Ile results in a mass shift that makes these two residues virtually indistinguishable from Arg [99]. Therefore, we filtered for only arginylation localized on the amino acids D or E, anywhere in the peptide sequence, except the N and C-terminal. Following these criteria, a total of 61 peptides, from 21 unique sequences were identified with potential arginylation (Supplementary Table S7).

We next performed a “closed” database search with +156.1011 and +128.1011 as variable modification in the N-terminal and

+156.1011 in specific amino acids such as D and E to confirm these modifications and identify the exact amino acid sites using a controlled database search. We identified 321 peptides with +156.1011 and 291 peptides with +128.1011 localized to the N-terminal (Supplementary Table S8). The overlap of unique peptide sequence identified using the “open” database search and “close” database search was 73 for the +156.1011 modification (Supplementary Fig. S7C) and 72 for the +128.1011 modifications in the N-terminal. Considering the total of unique sequence identified using the “open”



**Fig. 5.** Top 40 most abundant modification from  $-200$  to  $+200$  Da and their distribution over the twenty aminoacid are shown as heat map using the percentage of each modification per aminoacid.

database search, the overlap coverage 86% (73/85) and 90% (72/80) of +156.1011 and +128.1011 modifications, respectively. To confirm the presence of transpeptidation, we investigated peptides modified with delta mass shift of ( 156.101) and ( 128.101). Interestingly, the majority of the peptides modified with 156 contained arginine at the C-terminus, while the ones modified with 128 contained lysine at the C-terminal position (Supplementary Table S4). The localization of these modifications was at the peptide C-terminus, indicating that after trypsin cleavage these peptides could be the donor of arginine and lysine for the transpeptidation reaction. Furthermore, we also performed “closed” database search using the modification of +156.1011 on the amino acids D and E and we compared the identification with the modified peptides retrieved by Byonic. We filtered only the peptides identified with score localization > 0.9, resulting in 245 redundant peptides, from 191 unique sequence (Supplementary Table S9). The overlap between the 22 unique sequence identified using “open” database search with the “closed” database search was 10 peptides (45%) (Supplementary Fig. S7D). Representative spectra of peptides identified with the +156.1 modification in the internal amino acids D or E are shown in Supplementary Fig. S8. These identified peptides pointed towards the presence of arginylated peptides which could be added post-translationally in a process called protein arginylation [100]. Recently it was reported that Ate1 arginyltransferase had no significant activity toward non-canonical N-terminal or internal aspartic and glutamic acid residues [101]. These data were obtained using an *in vitro* system on 11-residues immobilized peptides array and pulse-chase arginylation

reaction raising concerns towards the mass spectrometry-based analyses. More studies are needed to validate the site-specific identifications obtained so far.

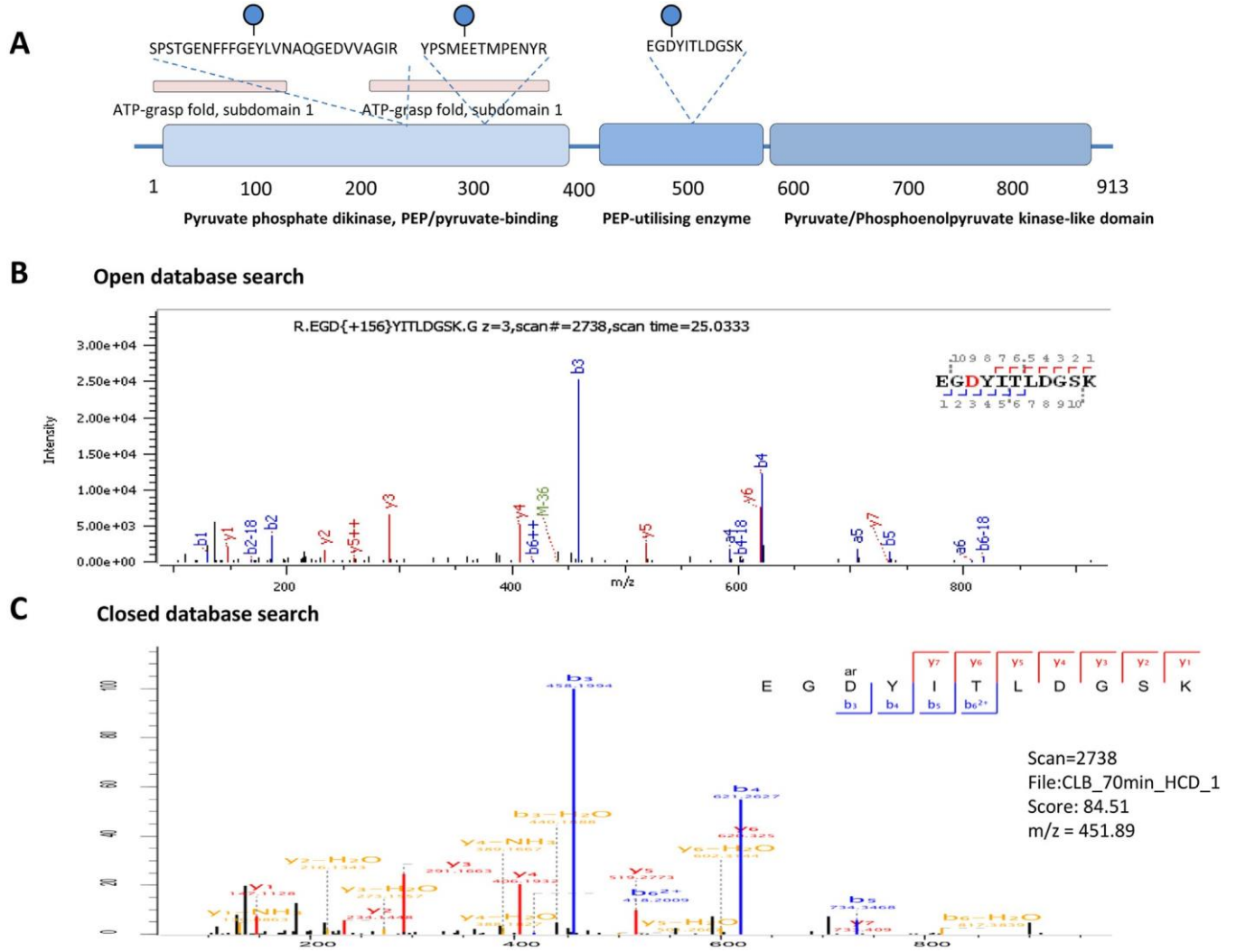
#### 1.10. Protein arginylation in trypanosomatids

Protein arginylation is a post-translational modification catalyzed by the addition of arginine to proteins in a ribosome-free

and tRNA-dependent reaction [102]. This modification is catalyzed by the arginyl-tRNA protein transferase (Ate1) enzyme [102] [103]. The enzyme is present in all eukaryotes from yeast to humans in single or multiple isoforms, respectively [97,104]. Initially, Ate1 was known to arginylate N-terminally exposed aspartic and glutamic acid [105,106], but recently it was shown that this modification can occur on various positions in the protein and on various amino acids side chains such as aspartic and glutamic acid [98,107]. Initial studies recognized the importance of protein arginylation related to the N-end rule of protein degradation [108]. Indeed, the addition of arginine, a primary destabilizing N-terminal residue, to proteins N-terminus regulates their *in vivo* half-life since it is recognized by ubiquitin ligase that lead to protein ubiquitination and degradation. This mechanism shows a direct interplay between protein arginylation and protein degradation. Nowadays, protein arginylation is recognized as a central biological pacemaker involved in a diverse array of physiological processes such as embryogenesis, aging [109], angiogenesis, response to stress [110] and brain physiology [109]. Deletion of the arginyltransferase enzyme has been shown to be fatal for mouse embryo and lead to severe pathological complications such as cardiovascular diseases [111], thrombosis [112], neurodegeneration [113] and tumor metastasis [114]. Due to that, protein arginylation is nowadays recognized as a widely present and functional post-translational modification.

The identification of Ate1 substrates has been lacking behind [115,116] until mass spectrometry-based proteomic analyses have allowed the identification of, 43 [117] and over hundreds proteins

[98] in two seminal studies. The initial amino acid specificity described at the N-terminally exposed aspartic acid, glutamic acid and oxidized cysteine was revised since it was shown that arginylation could occur on any N-terminally exposed residue [117] and on aspartic and glutamic acid side chain [98,107]. Moreover, several nuclear proteins contained post-translationally added arginines in



**Fig. 6.** Protein arginylation in Pyruvate phosphate dikinase 1 (Gene name: ppdk2, Uniprot ID Q9GN79). (A) Localization of peptides with +156.1 modification across ppdk2 domains. The arginylation site in the peptide sequence are shown in blue circles. Representative spectra for peptide EGDYITLDGSK with +156 modification in D3 retrieved by open database search using Byonic (B) and closed database search using Maxquant (C). The same scan number from the same file was identified using the different database search analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the methylated and dimethylated form and were shown to regulate nuclear structure [118].

In our study, one of the arginylated proteins was pyruvate phosphate dikinase 1 (ppdk2) that had the highest count for  $O_{mass}=+156.1$  in the amino acids E and D retrieved by the “closed” database search. We demonstrated the distribution of peptides identified with  $+156.1011$  across the annotated domains for ppdk2, according to InterPro database (Fig. 6a). Two peptides, containing the modification  $+156.1$  on glutamic acid, were identified in the PEP/pyruvate-binding domain (“closed” database search); one peptide containing the modification  $+156.1$  was identified using both closed and open database search and was localized in the PEP-utilizing enzyme. The role of these potential arginylations in ppdk2 on the activity and function is not known, but it is the first time, to our knowledge, that these modifications are identified in

*T. cruzi*. An example of annotated spectra retrieved by Byonic and MaxQuant for peptide EGDYITLDGS with  $O_{mass}$  modification of

$+156.1$  in the amino acid E are provided to demonstrate the confidence of identification of these modified peptides (Fig. 6b and c).

The addition of amino acids to proteins in a tRNA-dependent and protein translation-independent manner has been described

in eukaryotes and prokaryotes. In eukaryotes the atel1 enzyme is responsible for the post-translationally addition of arginine. In prokaryotes leucine and phenylalanine are conjugated to protein N-terminus by the Leu/Phe-tRNA-protein transferase [119]. Although these enzymes share similar function, there is low similarity in the primary amino acid sequence. Moreover, these enzymes can have mixed characteristic such as the bacterial protein transferase (Bpt) from *Vibrio vulnificus* has a sequence similarity with the eukaryotic atel1 but has a mixed enzymatic function transferring leucine to aspartic and glutamic acid. On the other hand, the ATE-like aminoacyl-tRNA-protein transferase, termed ATEL1, from *Plasmodium falciparum*, has sequence similarity with the prokaryotic Leu/Phe-tRNA-protein transferase but transfers arginine, as the eukaryotic atel1 [119]. The ATEL1 enzyme has an N-Acyltransferase superfamily (NAT SF) domain. So far, no enzyme with this function has been described in Trypanosomatids. We performed a BLASTp analysis of the ATEL1 sequence (Q195H6 accession in the UniprotKB) and found high similarity with the hypothetical protein (Q4DBY5 UniprotKB, Tc00.1047053506977.40 gene name, *Trypanosoma cruzi* strain CL Brener), Supplementary Fig. S9. Phylogenetic distance analysis showed that this enzyme is conserved in all trypanosomatids from *Trypanosoma* to *Leishmania*, Fig. 7. Inter-



**Fig. 7.** Phylogenetic analysis of the AtelI enzyme from *Plasmodium falciparum*. The different parasite species are grouped based on their phylum. Details on the phylogenetic tree construction are reported in the Materials and Methods section.



pro domain search identified a leucyl/phenylalanyl-tRNA protein transferase unintegrated signature motif (data not shown). More experiments are underway to confirm the enzymatic activity of this protein.

## Conclusions

In this study, we present a deep mass spectrometry data analysis of *T. cruzi* proteins and PTMs. A proteogenomic approach allowed us to identify five novel DNA coding regions, their subcellular localization and predicted PTMs. Moreover, an innovative application of unrestricted database search to Trypanosomatids is shown. This analysis allowed us to map for the first time several unknown PTMs on *T. cruzi* proteins. These modifications derived from artefactual sample preparation and/or cellular biosynthesis processes such as arginine addition ( $\Delta\text{mass} = +156.1$ ). This modification can be due to transpeptidation and protein arginylation. Protein arginylation has been reported in several organisms such as *E. coli*, *Arabidopsis*

and humans but not in Trypanosomatids. In this study, we report the presence of protein arginylation in *T. cruzi*. It should be noted, that the addition of arginine has been also described as a sample preparation artifact due to transpeptidation by trypsin. Due to that, it is very important to discriminate between them to avoid false positives. The correct localization of this modification has been a matter of concern due to the fact that chemical artefacts can generate similar mass shifts and generate false positives. High mass accuracy combined with stringent cutoffs and manual validation are needed to confirm the presence of post-translationally added protein arginylation [99]. However, transpeptidation is a common trypsin-catalyzed artefact that occurs during sample preparation and it is independent on the level of mass accuracy or computational cutoffs [120]. An initial solution to this dilemma was shown using a combination of inhibitors of protein synthesis, *ate1* gene knockout and bidimensional electrophoresis focusing on protein spots that were significantly regulated [117]. It should be noted, that although this approach significantly minimizes potential arte-

1

2 facts, it does not remove transpeptidation reaction that occurs during the proteolytic step.  
3 To address the issue of transpeptida- tion it was proposed to use alternative proteolytic  
4 enzymes such as AspN and GluC for shorter incubation times and alkaline pH [120]. Taken  
5 together these data show the need of further studies to iden- tify, quantify and minimize  
6 transpeptidation reaction in common trypsin-based bottom up proteomic workflows and  
7 identify true protein arginylation sites.

8 Overall, this study shows how it is important to evaluate the presence of  
9 unknown/unassigned PTMs in a bottom up mass spectrometry-based proteomic analysis.  
10 Indeed, by using an unre- stricted database search we showed an improvement of more than  
11 30% in PSMs, Table 1, and protein identification, Supplementary Fig. S10. Beside the  
12 enormous improvement in MS/MS spectra assignment shown here, there are several  
13 spectra that remained unassigned calling for further improvements in data analysis.

14 Moreover, this manuscript represents the first unbiased PTMs discovery in  
15 Trypanosomatids using large-scale mass spectrometry-based proteomics. Certainly, latest  
16 mass spectrom- eters with higher acquisition speed, resolution and accuracy will improve  
17 the current dataset and move towards the discovery of novel PTMs in Trypanosomatids.  
18 An initial step would be to start with data already present in the literature or deposited in  
19 public repositories.

20 Several large-scale mass spectrometry-based proteomic anal- yses have been performed  
21 using *T. cruzi* parasite. These studies have revealed differential expression of proteins and  
22 PTMs dur- ing development across life stages and biotic or abiotic stimuli. However,  
23 different strains have been used through these stud- ies making a proper comparison a hard  
24 task. Due to that, it is envisaged that the scientific community, working on *Trypanosoma*  
25 *cruzi*, establishes a reference strain with a sequenced and anno- tated genome for initial  
26 studies to be further compared with other strains. Indeed, after basic research and initial  
27 studies on a ref- erence strain, other strains belonging to different DTUs should be  
28 investigated to understand the genotype-phenotype-pathotype linkage. Moreover, it is  
29 advisable that an international effort such as the well-established platform TriTrypDB  
30 ([http://tritrypdb.org/ tritrypdb/](http://tritrypdb.org/tritrypdb/)) focus on trying to centralize and share the raw data of  
31 Trypanosomatids Proteomics.

32

### 33 Acknowledgements

34

35 This work was supported by a generous grant from the VIL- LUM Foundation for a VILLUM  
36 Center for Bioanalytical Sciences at the University of Southern Denmark (MRL). GP supported  
37 by CNPQ (441878/2014-8) and FAPESP (2014/06863-3). RK is supported by FAPESP  
38 (2015/02866-0) and GO is supported by Capes. Maria Julia

39 M. Alves from the IQ, USP, is acknowledged for suggestions during the proofreading of the manuscript.

40

### 41 Appendix A. Supplementary data

42

43 Supplementary data associated with this article can be found, in the online version, at  
44 <http://dx.doi.org/10.1016/j.ijms.2016.11.020>.

45

## References

- [1] A. Rassi Jr., A. Rassi, J.A. Marin-Neto, Chagas disease, *Lancet* (London, England) 375 (9723) (2010) 1388–1402.
- [2] WHO, Chagas Disease (*American trypanosomiasis*), World Health Organization, 2016 (2016 [cited 2016]).
- [3] J.R. Coura, S.L. De Castro, A critical review on chagas disease chemotherapy, *Mem. Inst. Oswaldo Cruz* (2002) 3–24.
- [4] B. Zingales, et al., A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI, *Mem. Inst. Oswaldo Cruz* 104 (7) (2009) 1051–1054.
- [5] Pinto, et al., TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification, *Infect. Genet. Evol.* 12 (6) (2012) 1328–1332.
- [6] L. Lima, et al., Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tebat as an independent DTU (discrete typing unit), *Acta Trop.* 151 (2015) 166–177.
- [7] P.N.B. Lanura, et al., Caracterização de cepas do *Trypanosoma cruzi* isoladas de doentes nos quais foi realizado transplante de coração, *Rev. Soc. Bras. Med. Trop.* 28 (1995) 351–356.
- [8] B. Zingales, et al., The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications, *Infect. Genet. Evol.* (2012) 240–253.
- [9] J.M. Burgos, et al., Differential distribution of genes encoding the virulence factor trans-sialidase along *Trypanosoma cruzi* discrete typing units, *PLoS One* 8 (3) (2013).
- [10] B. Zingales, *Trypanosoma cruzi*: um parasita, dois parasitas ou vários parasitas da doença de chagas? *Revista da Biologia* 6b (2011) 44–48.
- [11] J.C. Carranza, et al., *Trypanosoma cruzi* maxicircle heterogeneity in Chagas disease patients from Brazil, *Int. J. Parasitol.* 39 (9) (2009) 963–973.
- [12] N.M. El-Sayed, et al., Comparative genomics of trypanosomatid parasitic protozoa, *Science* (New York, N.Y.) 309 (5733) (2005) 404–409.
- [13] N.M. El-Sayed, et al., The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease, *Science* (New York, N.Y.) 309 (5733) (2005) 409–415.
- [14] O. Franzen, et al., Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi* marinkellei, *BMC Genom.* 13 (2012) 531.
- [15] E.C. Grisard, et al., *Trypanosoma cruzi* Clone Dm28c Draft Genome Sequence, *Genom. Announc.* 2 (1) (2014) 2–3.
- [16] A.P. Jackson, et al., The genome sequence of *Trypanosoma brucei* gambiense, causative agent of chronic human African Trypanosomiasis, *PLoS Negl. Trop. Dis.* 4 (4) (2010).
- [17] A.C. Ivens, et al., The genome of the kinetoplastid parasite: leishmania major, *Science* 309 (5733) (2005) 436–442.
- [18] J.D. Jaffe, H.C. Berg, G.M. Church, Proteogenomic mapping as a complementary method to perform genome annotation, *Proteomics* 4 (1) (2004) 59–77.
- [19] C. Ansong, et al., Proteogenomics: needs and roles to be filled by proteomics in genome annotation, *Brief. Funct. Genomic. Proteomic* 7 (1) (2008) 50–62.
- [20] J. Armengaud, A perfect genome annotation is within reach with the proteomics and genomics alliance, *Curr. Opin. Microbiol.* (2009) 292–300.
- [21] J.R. Yates, J.K. Eng, A.L. McCormack, Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases, *Anal. Chem.* 67 (18) (1995) 3202–3210.
- [22] N. Castellana, V. Bafna, Proteogenomics to discover the full coding content of genomes: A computational perspective, *J. Proteomics.* (2010) 2124–2135.
- [23] A.I. Nesvizhskii, Proteogenomics: concepts, applications and computational strategies, *Nat. Methods* 11 (11) (2014) 1114–1125.
- [24] S. Tanner, et al., Improving gene annotation using peptide mass spectrometry, *Genome Res.* 17 (2) (2007) 231–239.
- [25] E. Brunner, et al., A high-quality catalog of the *Drosophila melanogaster* proteome, *Nat. Biotechnol.* 25 (5) (2007) 576–583.
- [26] N. Gupta, et al., Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation, *Genome Res.* 17 (9) (2007) 1362–1377.
- [27] J.A. Atwood 3rd, et al., The *Trypanosoma cruzi* proteome, *Science* 309 (5733) (2005) 473–476.
- [28] J. Paba, et al., Proteomic analysis of the human pathogen *Trypanosoma cruzi*, *Proteomics* 4 (4) (2004) 1052–1059.
- [29] J.A. Atwood, et al., The *Trypanosoma cruzi* proteome, *Science* (New York, N.Y.) 309 (5733) (2005) 473–476.
- [30] A. Parodi-Talice, et al., Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis, *J. Mass Spectrom.* (2007).
- [31] S.A. Kikuchi, et al., Proteomic analysis of two *Trypanosoma cruzi* zymodeme 3 strains, *Exp. Parasitol.* 126 (4) (2010) 540–551.
- [32] D. Pérez-Morales, et al., Proteomic analysis of *Trypanosoma cruzi* epimastigotes subjected to heat shock, *J. Biomed. Biotechnol.* (2012) 2012.
- [33] G.V.F. Brunoro, et al., Reevaluating the *Trypanosoma cruzi* proteomic map: the shotgun description of bloodstream trypomastigotes, *J. Proteomics* 115 (2015) 58–65.
- [34] L.M. de Godoy, et al., Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis, *Proteomics* 12 (17) (2012) 2694–2703.
- [35] R.M. Queiroz, et al., Quantitative proteomic and phosphoproteomic analysis of *Trypanosoma cruzi* amastigogenesis, *Mol. Cell. Proteomics* 13 (12) (2014) 3457–3472.
- [36] E. Bayer-Santos, et al., Proteomic analysis of *Trypanosoma cruzi* secretome: characterization of two populations of extracellular vesicles and soluble proteins, *J. Proteome Res.* 12 (2) (2013) 883–897.
- [37] M. Pereira, et al., Down regulation of NO signaling in *Trypanosoma cruzi* upon parasite-extracellular matrix interaction: changes in protein modification by nitrosylation and nitration, *PLoS Negl. Trop. Dis.* 9 (4) (2015) e0003683.
- [38] F.K. Marchini, et al., Profiling the *Trypanosoma cruzi* phosphoproteome, *PLoS One* 6 (9) (2011) e25381.
- [39] E.S. Nakayasu, et al., GPIomics: global analysis of glycosylphosphatidylinositol-anchored molecules of *Trypanosoma cruzi*, *Mol. Syst. Biol.* 5 (2009) 261.
- [40] M.J. Alves, et al., Comprehensive glycoproteomic analysis of the epimastigote and trypomastigote stages of *Trypanosoma cruzi*, *J. Proteomics* (2016).
- [41] M.R. Larsen, et al., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns, *Mol. Cell. Proteomics* 4 (7) (2005) 873–886.
- [42] G. Palmisano, et al., Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry, *Nat. Protoc.* 5 (12) (2010) 1974–1982.
- [43] M. Mann, M. Wilm, Error-tolerant identification of peptides in sequence databases by peptide sequence tags, *Anal. Chem.* 66 (24) (1994) 4390–4399.
- [44] D.M. Creasy, J.S. Cottrell, Error tolerant searching of uninterpreted tandem mass spectrometry data, *Proteomics* 2 (10) (2002) 1426–1434.
- [45] S. Kim, et al., Spectral dictionaries: integrating de novo peptide sequencing with database search of tandem mass spectra, *Mol. Cell. Proteomics* 8 (1) (2009) 53–69.
- [46] M. Bern, Y. Cai, D. Goldberg, Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry, *Anal. Chem.* 79 (4) (2007) 1393–1400.
- [47] J.M. Chick, et al., A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides, *Nat. Biotechnol.* 33 (7) (2015) 743–749.
- [48] E.P. Camargo, Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of metacyclic trypanosomes in liquid media, *Rev. Inst. Med. Trop. Sao Paulo* 6 (1964) 93–100.
- [49] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (12) (2008) 1367–1372.
- [50] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *J. Am. Soc. Mass Spectrom.* 5 (11) (1994) 976–989.
- [51] V. Dorfer, et al., MS Amanda: a universal identification algorithm optimized for high accuracy tandem mass spectra, *J. Proteome Res.* 13 (8) (2014) 3679–3684.
- [52] J. Cox, et al., Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (4) (2011) 1794–1805.
- [53] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (12) (2008) 1367–1372.
- [54] B. Schwahnhauser, et al., Global quantification of mammalian gene expression control, *Nature* 473 (7347) (2011) 337–342.
- [55] M. Goujon, et al., A new bioinformatics analysis tools framework at EMBL-EBI, *Nucleic Acids Res.* 38 (Web Server issue) (2010) W695–W699.
- [56] Z. Zhang, et al., A greedy algorithm for aligning DNA sequences, *J. Comput. Biol.* 7 (1–2) (2000) 203–214.
- [57] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* 26 (6) (2010) 841–842.
- [58] R.C. Edgar, Quality measures for protein alignment benchmarks, *Nucleic Acids Res.* (2010).
- [59] M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Mol. Biol. Evol.* 27 (2) (2010) 221–224.
- [60] M.W. Bern, Y.J. Kil, Two-dimensional target decoy strategy for shotgun proteomics, *J. Proteome Res.* 10 (12) (2011) 5296–5301.
- [61] A. Marchler-Bauer, et al., CDD: NCBI's conserved domain database, *Nucleic Acids Res.* 43 (D1) (2015) D222–D226.
- [62] C.J.A. Sigrist, et al., New and continuing developments at PROSITE, *Nucleic Acids Res.* 41 (D1) (2013).
- [63] T.N. Petersen, et al., SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods* 8 (10) (2011) 785–786.
- [64] J.D. Bendtsen, et al., Feature-based prediction of non-classical and leaderless protein secretion, *Protein Eng. Des. Sel.* 17 (4) (2004) 349–356.
- [65] Y. Xue, et al., GPS: a comprehensive www server for phosphorylation sites prediction, *Nucleic Acids Res.* 33 (Suppl. 2) (2005).

- [66] A. Li, et al., Prediction of nepsilon-acetylation on internal lysines implemented in Bayesian discriminant method, *Biochem. Biophys. Res. Commun.* 350 (4) (2006) 818–824.
- [67] Q. Zhao, et al., GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs, *Nucleic Acids Res.* 42 (W1) (2014).
- [68] Z. Liu, et al., GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins, *Mol. Biosyst.* 7 (10) (2011) 2737–2740.
- [69] Z. Pan, et al., Systematic analysis of the in situ crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues, *Sci. Rep.* 4 (2014) 7331.
- [70] Y. Xue, et al., GPS-SNO: computational prediction of protein s-nitrosylation sites with a modified GPS algorithm, *PLoS One* 5 (6) (2010).
- [71] Y. Xie, et al., GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites, *Sci. Rep.* 6 (2016) 28249.
- [72] R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function, *Pac. Symp. Biocomput.* (2002) 310–322.
- [73] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics (Oxford, England)* 30 (9) (2014) 1312–1313.
- [74] Stöver, K.F. Müller, TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses, *BMC Bioinf.* 11 (1) (2010) 1–9.
- [75] D.H. Huson, et al., Dendroscope: an interactive viewer for large phylogenetic trees, *BMC Bioinf.* 8 (1) (2007) 1–6.
- [76] D.R. Stein, et al., High pH reversed-phase chromatography as a superior fractionation scheme compared to off-gel isoelectric focusing for complex proteome analysis, *Proteomics* 13 (20) (2013) 2956–2966.
- [77] F. Yang, et al., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis, *Expert Rev. Proteomics* 9 (2) (2012) 129–134.
- [78] A.M. Nichols, F.M. White, Manual validation of peptide sequence and sites of tyrosine phosphorylation from MS/MS spectra, *Methods Mol. Biol.* 492 (2009) 143–160.
- [79] T. Hunter, Tyrosine phosphorylation: thirty years and counting, *Curr. Opin. Cell Biol.* 21 (2) (2009) 140–146.
- [80] C. Doerig, Protein kinases as targets for anti-parasitic chemotherapy, *Biochim. Biophys. Acta* (2004).
- [81] F. Canduri, et al., Protein kinases as targets for antiparasitic chemotherapy drugs, *Curr. Drug Targets* 8 (3) (2007) 389–398.
- [82] N.S. Moretti, et al., Characterization of *Trypanosoma cruzi* Sirtuins as possible drug targets for chagas disease, *Antimicrob. Agents Chemother.* 59 (8) (2015) 4669–4679.
- [83] J.P. da Cunha, et al., Post-translational modifications of *Trypanosoma cruzi* histone H4, *Mol. Biochem. Parasitol.* 150 (2) (2006) 268–277.
- [84] J.C. Bayona, et al., SUMOylation pathway in *Trypanosoma cruzi*: functional characterization and proteomic analysis of target proteins, *Mol. Cell. Proteomics: MCP* 10 (2011) (p. M110.007369–M110.007369).
- [85] M.J. Pearce, et al., Ubiquitin-like protein involved in the proteasome pathway of mycobacterium tuberculosis, *Science* 322 (5904) (2008) 1104–1107.
- [86] L. Piacenza, et al., Fighting the oxidative assault: the *Trypanosoma cruzi* journey to infection, *Curr. Opin. Microbiol.* 12 (4) (2009) 415–421.
- [87] M. Dhiman, et al., Enhanced nitrosative stress during *Trypanosoma cruzi* infection causes nitrotyrosine modification of host proteins: implications in Chagas' disease, *Am. J. Pathol.* 173 (3) (2008) 728–740.
- [88] T. Furuya, et al., A novel phosphatidylinositol-phospholipase C of *Trypanosoma cruzi* that is lipid modified and activated during trypomastigote to amastigote differentiation, *J. Biol. Chem.* 275 (9) (2000) 6428–6438.
- [89] A.S. Hebert, et al., The one hour yeast proteome, *Mol. Cell. Proteomics* 13 (1) (2014) 339–347.
- [90] R.J. Chalkley, et al., In-depth analysis of tandem mass spectrometry data from disparate instrument types, *Mol. Cell. Proteomics* 7 (12) (2008) 2386–2398.
- [91] D.L. Swaney, C.D. Wenger, J.J. Coon, Value of using multiple proteases for large-scale mass spectrometry-based proteomics, *J. Proteome Res.* 9 (3) (2010) 1323–1329.
- [92] P. Giansanti, et al., An augmented multiple-protease-based human phosphopeptide atlas, *Cell Rep.* 11 (11) (2015) 1834–1843.
- [93] B. Thiede, et al., Analysis of missed cleavage sites: tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion, *Rapid Commun. Mass Spectrom.* 14 (6) (2000) 496–502.
- [94] J.M. Burkhardt, et al., Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics, *J. Proteomics* 75 (4) (2012) 1454–1462.
- [95] P. Picotti, R. Aebersold, B. Dörmann, The implications of proteolytic background for shotgun proteomics, *Mol. Cell. Proteomics* 6 (9) (2007) 1589–1598.
- [96] H. Schaefer, et al., Tryptic transpeptidation products observed in proteome analysis by liquid chromatography-tandem mass spectrometry, *Proteomics* 5 (4) (2005) 846–852.
- [97] R. Rai, A. Kashina, Identification of mammalian arginyltransferases that modify a specific subset of protein substrates, *Proc. Natl. Acad. Sci. U. S. A.* 102 (29) (2005) 10123–10128.
- [98] J. Wang, et al., Arginyltransferase ATE1 catalyzes midchain arginylation of proteins at side chain carboxylates in vivo, *Chem. Biol.* 21 (3) (2014) 331–337.
- [99] T. Xu, et al., Identification of N-terminally arginylated proteins and peptides by mass spectrometry, *Nat. Protoc.* 4 (3) (2009) 325–332.
- [100] A. Kashina, Protein arginylation: a global biological regulator that targets actin cytoskeleton and the muscle, *Anat. Rec. (Hoboken)* 297 (9) (2014) 1630–1636.
- [101] B. Wadas, et al., Analyzing N-terminal arginylation through the use of peptide arrays and degradation assays, *J. Biol. Chem.* 291 (40) (2016) 20976–20992.
- [102] A. Kaji, H. Kaji, G.D. Novelli, A soluble amino acid incorporating system, *Biochem. Biophys. Res. Commun.* 10 (5) (1963) 406–409.
- [103] E. Balzi, et al., Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of *Saccharomyces cerevisiae*, *J. Biol. Chem.* 265 (13) (1990) 7464–7471.
- [104] Y.T. Kwon, A.S. Kashina, A. Varshavsky, Alternative splicing results in differential expression: activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway, *Mol. Cell. Biol.* 19 (1) (1999) 182–193.
- [105] R.L. Soffer, Enzymatic modification of proteins. 4. Arginylation of bovine thyroglobulin, *J. Biol. Chem.* 246 (5) (1971) 1481–1484.
- [106] R.L. Soffer, H. Horinishi, Enzymic modification of proteins. I. General characteristics of the arginine-transfer reaction in rabbit liver cytoplasm, *J. Mol. Biol.* 43 (1) (1969) 163–175.
- [107] E. Eriste, et al., A novel form of neurotensin post-translationally modified by arginylation, *J. Biol. Chem.* 280 (42) (2005) 35089–35097.
- [108] A. Bachmair, D. Finley, A. Varshavsky, In vivo half-life of a protein is a function of its amino-terminal residue, *Science* 234 (4773) (1986) 179–186.
- [109] K.D. Lamon, H. Kaji, Arginyl-tRNA transferase activity as a marker of cellular aging in peripheral rat tissues, *Exp. Gerontol.* 15 (1) (1980) 53–64.
- [110] K.D. Lamon, W.H. Vogel, H. Kaji, Stress-induced increases in rat brain arginyl-tRNA transferase activity, *Brain Res.* 190 (1) (1980) 285–287.
- [111] Y.T. Kwon, et al., An essential role of N-terminal arginylation in cardiovascular development, *Science* 297 (5578) (2002) 96–99.
- [112] L. Lian, et al., Loss of ATE1-mediated arginylation leads to impaired platelet myosin phosphorylation: clot retraction, and in vivo thrombosis formation, *Haematologica* 99 (3) (2014) 554–560.
- [113] C.S. Brower, K.I. Piatkov, A. Varshavsky, Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway, *Mol. Cell* 50 (2) (2013) 161–171.
- [114] Pavlyk, et al., Arginine deprivation affects glioblastoma cell adhesion: invasiveness and actin cytoskeleton organization by impairment of beta-actin arginylation, *Amino Acids* 47 (1) (2015) 199–212.
- [115] I.V. Davydov, A. Varshavsky, RGS4 is arginylated and degraded by the N-end rule pathway in vitro, *J. Biol. Chem.* 275 (30) (2000) 22931–22941.
- [116] J. Kopitz, B. Rist, P. Böhley, Post-translational arginylation of ornithine decarboxylase from rat hepatocytes, *Biochem. J.* 267 (2) (1990) 343–348.
- [117] C.C. et al. Wong, Global analysis of posttranslational protein arginylation, *PLoS Biol.* 5 (10) (2007) e258.
- [118] S. Saha, et al., Arginylation and methylation double up to regulate nuclear proteins and nuclear architecture in vivo, *Chem. Biol.* 18 (11) (2011) 1369–1378.
- [119] J.W. Tobias, et al., The N-end rule in bacteria, *Science* 254 (5036) (1991) 1374–1377.
- [120] S. Hara, R. Rosenfeld, H.S. Lu, Preventing the generation of artifacts during peptide map analysis of recombinant human insulin-like growth factor-I, *Anal. Biochem.* 243 (1) (1996) 74–79.

139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226

227

## APÊNDICE B

228 **ARTIGO SUBMETIDO**

229

230 **Title:** Development of a *Trypanosoma cruzi* Strain Typing Assay using MS2 peptide spectral libraries  
231 (Tc-STAMS2).

232 **Running title:** Tc-STAMS2 as a novel tool for *T.cruzi* DTUs discrimination

233

234 Gilberto Santos de Oliveira<sup>1</sup>, Rebeca Kawahara<sup>1</sup>, Livia Rosa-Fernandes<sup>2</sup>, Carla Cristi Avila<sup>1</sup>, Marta M.

235 G. Teixeira<sup>1</sup>, Martin R. Larsen<sup>2</sup> and Giuseppe Palmisano<sup>1\*</sup>

236

237 **Affiliations:**

238 <sup>1</sup>Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo,

239 Brazil

240 <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense,

241 Denmark

242

243 \* To whom correspondence should be addressed: GlycoProteomics Laboratory, Department of

244 Parasitology, Institute of Biomedical Sciences, University of São Paulo, Avenida Lineu Prestes 1374,

245 CEP: 05508-000, São Paulo, Brazil. Email: palmisano.gp@usp.br; palmisano.gp@gmail.com

246

247 **Abstract**

248 Background: Chagas disease also known as American trypanosomiasis is caused by the protozoan

249 *Trypanosoma cruzi*. Over the last 30 years, Chagas disease has expanded from a neglected parasitic

250 infection of the rural population to an urbanized chronic disease, becoming a potentially emergent

251 global health problem. *T. cruzi* strains were assigned to seven genetic groups (TcI-TcVI and TcBat),

252 named discrete typing units (DTUs), which represent a set of isolates that differ in virulence,

253 pathogenicity and immunological features. Indeed, diverse clinical manifestations (from

254 asymptomatic to highly severe disease) have been attempted to be related to *T.cruzi* genetic variability.

255 Due to that, several DTU typing methods have been introduced. Each method has its own advantages

256 and drawbacks such as high complexity and analysis time and all of them are based on genetic

257 signatures. Recently, Shao W. et al.(Shao *et al.*, 2015) discriminated bacterial strains using a peptide

258 identification-free, genome sequence-independent shotgun proteomics workflow. Here, we aimed to

259 develop a *Trypanosoma cruzi* Strain Typing Assay using MS/MS peptide spectral libraries, named Tc-

260 STAMS2.

261 **Methods/Principal findings:** The Tc-STAMS2 method uses shotgun proteomics combined with spectral  
262 library search to assign and discriminate *T. cruzi* strains independently on the genome knowledge.  
263 The method is based on the construction of a library of MS/MS peptide spectra built using genotyped  
264 *T. cruzi* reference strains. For identification, the MS/MS peptide spectra of unknown *T. cruzi* cells are  
265 identified using the spectral matching algorithm SpectraST. The Tc-STAMS2 method allowed correct  
266 identification of all DTUs with high confidence. The method was robust towards different sample  
267 preparations, length of chromatographic gradients and fragmentation techniques. Moreover, a pilot  
268 inter-laboratory study showed the applicability to different MS platforms.

269 **Conclusions and significance:** This is the first study that develops a MS-based platform for *T. cruzi* strain  
270 typing. Indeed, the Tc-STAMS2 method allows *T. cruzi* strain typing using MS/MS spectra as  
271 discriminatory features and allows the differentiation of TcI-TcVI DTUs. Similar to genomic-based  
272 strategies, the Tc-STAMS2 method allows identification of strains within DTUs. Its robustness towards  
273 different experimental and biological variables makes it a valuable complementary strategy to the  
274 current *T. cruzi* genotyping assays. Moreover, this method can be used to identify DTU-specific  
275 features correlated with the strain phenotype.

276 **Author summary:** Chagas disease is one of the most important neglected diseases with an estimated  
277 number of 12 million infected individuals, the majority living in Central and South America. The  
278 *Trypanosoma cruzi* (*T.cruzi*) protozoan parasite is the etiological agent of Chagas disease. *T.cruzi* is  
279 highly genetically diverse and a new nomenclature assigned each strain to seven genetic groups (TcI-  
280 TcVI and Tcbat), named Discrete Typing Units (DTUs), based on their biochemical, immunological and  
281 phenotypical characteristics. *T.cruzi* DTUs have been correlated to diverse clinical outcomes  
282 highlighting the importance of molecular epidemiological screens. Despite the development of *T.cruzi*  
283 typing methods based on genetic signatures, each method presenting its own advantages and  
284 challenges. The work presented here shows the application of mass spectrometry for *Trypanosoma*  
285 *cruzi* Strain Typing Assay using MS<sup>2</sup> peptide spectral libraries (Tc-STAMS2). The novelty of the method  
286 is based on the use of peptide fragmentation spectra as strain-specific fingerprints to classify and  
287 identify DTUs. Initially, a spectra library is generated from characterized *T.cruzi* strains. The library is  
288 subsequently inspected using MS/MS spectra from unknown strains and confidently assigned to a  
289 specific strain in an automated and computationally-driven approach. The Tc-STAMS2 method was  
290 challenged to test several variables such as sample type and preparation, instrument setup and  
291 identification platform. Tc-STAMS2 provided high confidence and robustness in *T.cruzi* strain typing.  
292 The Tc-STAMS2 method represents a proof-of-concept of a complementary strategy to the current

293 DNA-based *T. cruzi* genotyping methods. Moreover, the method allows the identification of strain-  
294 specific features that could be related to the biology of *T. cruzi* strains and their clinical outcomes.

295 **Keywords:** *Trypanosoma cruzi*; Discrete Typing Units (DTUs); Mass spectrometry; Spectral matching;  
296 Strain typing methods.

297

## 298 Introduction

299 Chagas disease also known as American trypanosomiasis affects around 12 million people  
300 especially in Latin America. The etiologic agent of Chagas diseases is the protozoan *Trypanosoma cruzi*  
301 (Who, 2015) that infects several mammalian hosts and is primarily transmitted through the  
302 contamination with feces of triatomine bugs. Besides, congenital, blood transfusions, transplants and  
303 ingestion of contaminated foods represent other ways of infection (Rassi a Jr, 2010). Chagas disease  
304 is characterized by an acute and chronic phase. The acute phase lasts a few weeks and present mild  
305 symptoms such as fever and swelling around the site of infection. The chronic phase is in general  
306 lifelong and asymptomatic. However, 20-30% of patients develop cardiac or gastrointestinal  
307 complications(Rassi a Jr, 2010).

308 *T. cruzi* is highly genetically diverse. In order to standardize the nomenclature facilitating the  
309 communication among scientists, *T. cruzi* strains were divided into six (Tc I-VI) *discrete typing units*  
310 (DTUs) plus Tcbat, a novel strain associated with bats. Each group represents a set of isolates that are  
311 genetically similar and can be identified by common immunological, biochemical, pathological and  
312 molecular markers (Zingales *et al.*, 2009b). *T. cruzi* strains characterization is extremely important to  
313 understand different epidemiological and pathological characteristics such as geographical  
314 distribution and clinical outcomes (Zingales, 2011b). Several techniques have been introduced to  
315 improve the genotyping. In particular, the genetic diversity of *T. cruzi* was first recognized by  
316 multilocus enzyme electrophoresis (MLEE) and restriction analysis of kinetoplastid DNA minicircles  
317 (Henriksson *et al.*, 1990b; Macedo *et al.*, 1992b). Currently, the methods used for typing of strains of  
318 *T. cruzi* are based on polymorphism of the mini-exon gene (Spliced Leader) and the 24S $\alpha$  and 18S  
319 ribosomal RNA (Brisse *et al.*, 2001a). Other assays involve the analysis of complex electrophoretic  
320 patterns generated by restriction polymorphisms of PCR amplified genomic DNA (Tibayrenc *et al.*,  
321 1993b; Cosentino e Agüero, 2012a; Messenger *et al.*, 2015). These methods are able to discriminate  
322 *T. cruzi* strains, but they are work and time consuming and the interpretations of the results may be  
323 misleading. Multiplex Real-Time PCR Assay Using TaqMan Probes proved useful to determine DTUs of  
324 cultured *T. cruzi*, vector and blood samples from patients in acute infection (Cosentino e Agüero,  
325 2012a; Cura, Carolina I. *et al.*, 2015; Messenger *et al.*, 2015).

326 Proteomics methods using mass spectrometry have emerged as a powerful strategy to discriminate  
327 bacteria and have been established as a valuable alternative to DNA-based bacterial identification.  
328 Indeed, MALDI-TOF MS generates an intact protein profile that is compared to a MS database of  
329 known species. This method is easy to perform, rapid and provide accurate results and has been  
330 introduced as routine in most hospitals (Singhal *et al.*, 2015). However, this method is currently not  
331 able to distinguish at the strain level. Shotgun sequencing of peptides, derived from enzymatically  
332 digested proteins, using LC-MS/MS in combination with database search of *in silico* digested proteins  
333 derived from publicly available protein sequence has been proposed as a method to discriminate  
334 bacterial strains such as *Helicobacter pylori* and *Yersinia persis* (Shao *et al.*, 2015). This approach has  
335 great discrimination power compared to MALDI-TOF MS since it allows the identification of thousands  
336 of peptides with higher dynamic range. However, this method suffers from lack of suitable databases  
337 since many species have not been sequenced. Recently, a novel method based on MS/MS spectral  
338 matching has been shown to identify the blood meal of ticks (*Ixodes scapularis*) (Onder *et al.*, 2013)  
339 and has also been applied to discriminate *E.coli* strains from different isolates (Shao *et al.*, 2015).

340 Few studies have explored the possibility of using mass spectrometry as a diagnostic tool for Chagas  
341 disease. Using mass spectrometry, Ndao M. *et al.* (Ndao, 2009) were able to identify full-length Apo1  
342 as a serum biomarker of chronic Chagas disease patients using a SELDI-TOF approach. Despite several  
343 proteomic studies to understand the molecular features of *T. cruzi* in different biological conditions  
344 (Atwood, J. A. *et al.*, 2005; Parodi-Talice *et al.*, 2007; De Godoy *et al.*, 2012; Perez-Morales *et al.*, 2012;  
345 Brunoro *et al.*, 2015). To date, there is no report on the use of mass spectrometry for assaying various  
346 *T. cruzi* strains.

347 A pioneer work developed a peptide identification-free shotgun proteomics workflow to trace the  
348 vertebrate host that a tick (*Ixodes scapularis*) was feeding on (Onder *et al.*, 2013). This workflow used  
349 MS/MS spectral matching. The same strategy, named UNID, was used to profile bacterial strains (Shao  
350 *et al.*, 2015). In this study, we developed a platform to discriminate *T. cruzi* strains using MS/MS  
351 peptide spectral matching (Tc-STAMS2). The method was able to discriminate *T. cruzi* strains from  
352 different origins with high sensitivity and accuracy. The method was robust towards sample  
353 preparation and instrumental parameters, such as peptide purification, chromatographic gradient  
354 time, peptide fragmentation techniques and MS instruments. The MS/MS spectra were subjected to  
355 database search. More than 4000 proteins were identified in the combined six DTUs strains analyzed.  
356 A total of 1096 proteins were differentially expressed between the six DTUs and multivariate analysis  
357 allowed the discrimination of *T. cruzi* strains using the quantitative MS signal. In conclusion, this study  
358 describes a mass spectrometry-based method to discriminate *T. cruzi* strains. The Tc-STAMS2 method



359 represents a proof-of-concept of an alternative strategy to DNA-based *T. cruzi* genotyping. Further  
360 studies are needed to show its applicability to biofluids in clinical isolates.

361

## 362 **Methods**

### 363 **1) *T. cruzi* cultures**

364 Epimastigote forms of *T. cruzi* cultivated in LIT (Liver Infusion Tryptose) supplemented with 10% fetal  
365 bovine serum (Camargo, E., 1964) at 28°C of exponential culture phase were employed in the present  
366 study. DTU classification of all *T. cruzi* strains were confirmed by sequencing (Lima, Espinosa-Alvarez,  
367 Pinto, *et al.*, 2015). Only validated DTUs strains were used in this study (**Table 1**).

#### 368 **1.1) Cell growth challenging test**

369 *T. cruzi* cells from the *Sylvio X10/1* DTU I strain were collected in the exponential and stationary phase  
370 and processed as described below. Two biological replicates for the stationary (St\_1 and St\_2) and  
371 exponential (Exp\_1 and Exp\_2) growth phase were analyzed.

### 372 **2) Sample preparation and nLC-MS/MS analysis**

#### 373 **2.1) Protein extraction and digestion**

374 Epimastigote forms ( $5 \times 10^8$ ) were washed three times in phosphate buffered saline (PBS), pH 7.2  
375 (8,000g for 10 minutes at room temperature), and was re-suspended in 400  $\mu$ L of lysis buffer (7M  
376 urea, 2M thiourea, 1 mM DTT and protease inhibitors (Amersham) and incubated under stirring for  
377 30 minutes to solubilize the proteins. Proteins were reduced with 10 mM DTT (DL-Dithiothreitol–  
378 Sigma-Aldrich), alkylated with 40 mM iodoacetamide (Sigma-Aldrich), digested with trypsin  
379 (Promega) in the ratio 1:50 ( $\mu$ g trypsin/ $\mu$ g protein) in 50 mM ammonium bicarbonate solution at 37°C  
380 overnight. The reaction was stopped with 1% formic acid (less than pH 3) and then the sample was  
381 desalted with C18 columns (StageTips). Four biological replicates were prepared for each DTU. Blind  
382 test samples A (DTU-III) and B (DTU-I) were prepared using a minimum of three biological replicates  
383 according to the protocol described above.

#### 384 **2.2) Sample preparation for the challenging tests**

385 Peptide desalting was performed in acid (acid 1 and acid 2) and basic (basic 1 and basic 2) conditions.  
386 In particular, for the acid purification tryptic peptides were acidified with 0.1% TFA (pH 3) and loaded

387 onto an acid-activated StageTip microcolumn before being eluted with 50% acetonitrile: 0.1% TFA.  
388 For the basic purification, tryptic peptides were dissolved in 0.1% ammonia water (pH 10) and loaded  
389 onto a base-activated StageTip before being eluted with 50% acetonitrile: 0.1% ammonia water. The  
390 eluted peptides were lyophilized and analyzed by liquid chromatography tandem mass spectrometry  
391 (LC-MS/MS).

### 392 2.3) Nano LC-MS/MS Analysis

393 Peptides were separated by Reprosil-Pur C18-AQ column (3 $\mu$ m; Dr. Maisch GmbH, Germany) using  
394 Easy nano-LC HPLC (Proxeon, Odense, Denmark). The HPLC gradient was 0-34% B solvent (A = 0.1%  
395 formic acid; B = 90% ACN, 0.1% formic acid) in 70 min at a flow of 250 nL/min. The MS analysis was  
396 performed using the LTQ-Orbitrap Velos (Thermo Scientific, Bremen, Germany). The mass range was  
397 400-1500  $m/z$  at a resolution of 30,000 at 400  $m/z$  for a target value of  $1e^6$  ions. For each MS scan,  
398 collision induced dissociation (CID) fragmentation was performed on the 20 most intense ions in the  
399 linear iontrap. The parameters for data acquisition were: activation time = 15 ms, normalized energy  
400 = 35, Q-activation = 0:25, exclusion = available with repeat count 1, exclusion duration = 30s and  
401 intensity threshold = 30.000, target ions =  $2e^4$  (Palmisano *et al.*, 2012b). All raw data have been  
402 submitted to PRIDE archive (<https://www.ebi.ac.uk/pride/archive/>).

### 403 2.4) Sample amounts, chromatographic gradients and MS/MS fragmentation types used for the 404 challenging test

405 The robustness of the Tc-STAMS2 method was tested using different parameters: 1) sample amounts,  
406 2) chromatographic gradients and 3) MS fragmentation techniques. For the different sample  
407 amounts, 0.5 $\mu$ g and 1 $\mu$ g, were loaded onto the analytical column before MS analysis and named low  
408 and high, respectively. The chromatographic elution time was set to 20, 70 and 130 min from 0-34%  
409 B solvent at 250nL/min. CID fragmentation was used to develop the Tc-STAMS2 method. Higher  
410 energy collision induced dissociation (HCD) was evaluated as alternative peptide fragmentation type  
411 on tryptic peptides separated on a 70min chromatographic gradient. For the HCD fragmentation,  
412 each MS scan was acquired at resolution of 30,000 FWHM followed by 7 MS/MS scan of the most  
413 intense ions with an activation time of 0.1 ms and normalized collision energy of 35. The spectral  
414 library for each DTU was developed on a LTQ-Orbitrap Velos Pro instrument (Thermo Fisher Scientific)  
415 located in the PR group, Department of Biochemistry and Molecular biology, University of Southern  
416 Denmark. All the other tests for testing and validating the method were performed on a LTQ-Orbitrap

417 Velos Normal at the Biomass mass spectrometry facility (São Paulo, Brazil).

### 418 3) Bioinformatics and statistical analyses

#### 419 3.1) MS/MS spectral library generation and spectral matching

420 MS/MS spectral library generation and spectral matching were performed using the SpectraST  
421 software (version 4.8) as previously described (Lam *et al.*, 2007; Onder *et al.*, 2013; Shao *et al.*, 2015).  
422 In particular, the LC-MS/MS acquired spectra were converted to an open format (mzXML) by  
423 MSconvert (Kessner *et al.*, 2008), forming part of the software suite offered by TPP (Trans-Proteomic  
424 Pipeline) (Deutsch *et al.*, 2010). SpectraST (version 4.8) was used to build the reference spectral library  
425 and perform MS/MS spectral matching (Lam *et al.*, 2007). The reference spectral library was  
426 generated with three raw files for each one of the six DTUs and one raw file for each DTU was  
427 compared against constructed reference library. The first step for the reference spectral library  
428 generation involves the application of a threshold at which MS/MS spectra originated from the same  
429 peptide precursor ion are combined to create the consensus spectrum. Moreover, low quality spectra  
430 are excluded from the library (Lam *et al.*, 2007). To determine the spectrum of similarity between the  
431 query spectrum and the reference library, SpectraST uses the Spectral Dataset Similarity (SDSS)  
432 function (Stein, S. E. e Scott, D. R., 1994; Onder *et al.*, 2013). In particular, the unique dot product  
433 SDSS, abbreviated as “score” along the text, was chosen as *T. cruzi* DTU strain discrimination function  
434 and reported (Onder *et al.*, 2013). The statistical confidence in the identification of the correct DTU is  
435 made by data bootstrap (Onder *et al.*, 2013). All the spectral matching experiments reported below  
436 had a bootstrap of 1 unless reported.

437 DiagnoProt software was used also to match the MS/MS spectra of unknown samples against a  
438 database of genotyped *T. cruzi* strains (Silva *et al.*, 2017b). Default parameters for creating the  
439 spectral database were used: similarity threshold 0.70, precursor tolerance 4.50; activation type CID;  
440 minimum number peaks: 50; minimum relativity intensity: 0.01; minimum retention time 10.00; bin  
441 offset 0.40; bin size: 1.0005; minimum bin  $m/z$  200.00; maximum bin  $m/z$  1700.00. The spectral  
442 database was used to match the identity of each unknown sample.

443

#### 444 3.2) Database searches

445 The raw LC-MS/MS files were analyzed using: Proteome Discoverer, MaxQuant and the  
446 TransProteomic Pipeline. Proteome Discoverer v2.1 (Thermo Scientific) was used with the *T. cruzi*  
447 database using Mascot and Sequest. The searches in the database were conducted with the following  
448 parameters: precursor mass tolerance of 20 ppm; MS/MS mass tolerance 0.5 Da (CID data). Trypsin

449 was selected as enzyme and carbamidomethyl cysteine as fixed modification. The variables  
450 modifications were oxidation of methionine and deamidation (NQ). Shared peptide sequences were  
451 grouped as grouped accessions proteins. The False Discovery Rates (FDR) was calculated using the  
452 algorithm Percolator with  $q$  equal or less than 0.01. Protein FDR was calculated in the Proteome  
453 Discoverer software and kept below 1%.

454 The raw files were also processed using the MaxQuant (Cox e Mann, 2008) version 1.2.7.429 and the  
455 MS/MS spectra were searched using the Andromeda search engine (Cox *et al.*, 2011) against the  
456 Uniprot *T. cruzi* Protein Database (release July 11, 2017; 51,738 entries). The initial maximal allowed  
457 mass tolerance was set to 20 ppm for precursor and then set to 4.5 ppm in the main search and to  
458 0.5 Da for fragment ions. Enzyme specificity was set to trypsin with a maximum of two missed  
459 cleavages. Carbamidomethylation of cysteine (57.02 Da) was set as a fixed modification, and  
460 oxidation of methionine (15.99 Da), deamidation (NQ) and protein N-terminal acetylation (42.01 Da)  
461 were selected as variable modifications. Bioinformatics analysis was performed using the software  
462 Perseus v.1.5.2.6 (Cox e Mann, 2008) available in the MaxQuant environment and reverse and  
463 contaminant entries were excluded from further analysis. Protein FDR was calculated in the  
464 MaxQuant software and kept below 1%. Label Free Quantification (LFQ) intensity values were  
465 considered to relatively compare the abundance of proteins present in the different DTUs.

466 TransProteomic Pipeline software suite was also used to search raw files converted to mzXML (Pedrioli,  
467 2010). The mzXML files were searched by the Comet search algorithm embedded into the TPP  
468 platform (Eng *et al.*, 2013). Peptide and protein FDR was estimated using the PeptideProphet and  
469 ProteinProphet algorithm embedded in the TransProteomic Pipeline (Keller *et al.*, 2002).  
470 Identifications with less than 1% FDR were kept.

471 Raw data from human placental tissue (Lee *et al.*, 2013) and *T. vivax* (Meta, BSF1 and EP1) (Jackson  
472 *et al.*, 2015) were obtained from the public MS spectra databank PRoteomics IDentifications (PRIDE)  
473 and used for the negative control test.

474

475

476

#### 477 **4) Statistical analyses**

478 Label-free quantified peptides/proteins were analysed by the Perseus software (Tyanova *et al.*, 2016).  
479 Significantly regulated features with a p value less than 0.05 corrected with the Bonferroni post-hoc  
480 test were used to cluster the different DTUs. Hierarchical clustering of significantly regulated  
481 proteins/peptides was performed using the Z-score calculation on the log<sub>2</sub> intensity values and it was

482 represented as a heat map. The Principal Component Analysis (PCA) was performed using the same  
483 procedure described above in Perseus software. In addition, for generation of Venn diagram we used  
484 Venn Diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/> Bioinformatics & Evolutionary  
485 Genomics).

486  
487

## 488 Results

489 Tc-STAMS2 strategy allows DTUs discrimination

490 In this study, the combination of mass spectrometry and computational approaches was used to  
491 develop a method for *T. cruzi* DTU discrimination, named Tc-STAMS2. A schematic overview of the *T.*  
492 *cruzi* DTUs identification using MS/MS spectra from tryptic peptides is summarized in **Fig. 1**. A  
493 reference spectral library was built using a total of 586513 unique MS/MS spectra of tryptic peptides  
494 derived from three raw MS files of each one of the six *T. cruzi* strains (Figure 1A). Each *T. cruzi* strain  
495 was processed and acquired in four biological replicates. MS/MS spectra acquired from three  
496 replicates of each DTU were used to build the reference spectral library using SpectraST software (Lam  
497 *et al.*, 2007). Following the construction of the reference MS/MS library, proteins from unknown *T.*  
498 *cruzi* strain samples were extracted and digested with trypsin before being analyzed by nLC-MS/MS  
499 (Figure 1B). A LC-MS profile of four replicates of DTU-I is reported in Supplementary Figure S1 and the  
500 Pearson correlation score indicates high similarity between the different runs. The LC-MS  
501 chromatographic profile of the tryptic peptides belonging to the six *T. cruzi* strains shows high  
502 similarity between the different DTUs, Supplementary Figure S2. MS/MS spectra from different DTUs  
503 were subjected to spectral matching comparison with the library using SpectraST software (Lam *et al.*,  
504 2008)). The identification was made by finding the reference library with the highest similarity to the  
505 sample spectral dataset, in this case, the different DTUs, to be tested. Unique dot product SDSS score  
506 was used to provide a quantitative similarity measure between two spectral datasets (Figure 1C)  
507 (Onder *et al.*, 2013). MS/MS spectra were searched against the *T. cruzi* proteome database as  
508 described below (Figure 1D).

509

510 Based on the similarity scores between the MS/MS spectra from an unknown sample with the mass  
511 spectral library, the Tc-STAMS2 was able to differentiate and accurately identify the different DTUs.  
512 We observed that the score values were between 0.75 to 0.86 for true matches and close to 0 in  
513 unmatched cases, as shown in **Table 2**.

514 The LC-MS chromatograms obtained for each replicate and for each DTU (**Supplementary Figure S1**  
515 **and S2**) showed high similarity; however, the developed method was capable of differentiating and

516 identifying each of them. In order to rule out the possibility that different growth phases could  
517 influence the assignment of the algorithm, *Sylvio X10/1* (DTU-I) epimastigotes in the exponential and  
518 stationary phase were collected. Independently of the growth phase, the algorithm was able to assign  
519 it to the correct DTU, Table 3. This demonstrates that the identification method is not affected by the  
520 phases of the parasite.

521

522 Moreover, the performance of the Tc-STAMS2 spectral matching approach were tested to correctly  
523 identify MS/MS data sets from: 1) a *T. cruzi* strain that is known to belong to DTU-VI (*CL14*) (Zingales  
524 *et al.*, 2009b) but was not included in the spectral library, 2) from a species phylogenetically related,  
525 such as *Trypanosoma vivax* and 3) from species with completely distant genome (ex: human, *E. coli*,  
526 mouse). For this analysis, a new MS/MS library was constructed, using the MS/MS spectra from the  
527 six DTUs, including *T. cruzi CL14* and *T. vivax* in metacyclic stage (meta1 e meta2). Firstly, MS/MS  
528 spectra from *T. cruzi* strains *CL14* were compared with the library and the similarity score matched to  
529 the *CL14* (score = 0.417). Interestingly, although the similarity scores of *CL14* with DTUs I to V were  
530 close to zero, the similarity between *CL14* and DTU-VI was comparatively high (score = 0.133),  
531 indicating that many spectra MS/MS of *CL14* are shared with DTU-VI (CL Brener), as shown in **Figure**  
532 **2**. It should be noted that the LC-MS chromatographic profiles of the *CL14* and CL Brener strains have  
533 very high similarity, **Supplementary Figure S3**. However, the Tc-STAMS2 was able to clearly differentiate  
534 between the two strains within the same DTU.

535 Subsequently, MS/MS spectra from three different life stages of *T. vivax*, Meta3 – metacyclic phase,  
536 BSF1 – bloodstream phase and EP1 – epimastigote were compared using the mass spectral library.  
537 Based on the similarity scores, the Tc-STAMS2 was able to correctly identify these samples to *T. vivax*,  
538 Figure 2.

539 In order to test the method with negative control, MS/MS spectra from unrelated *T. cruzi* species such  
540 as human, mouse and *E. coli* were compared to the library. For these samples, the similarity scores  
541 were close to zero, indicating that the similarity scores found between two MS/MS datasets is library-  
542 specific and not random, **Figure 2**.

543 Moreover, we also evaluated the ability of this strategy to provide correct identification from an  
544 independent sample (blind test), which was collected and processed at different days or under  
545 different conditions. In particular the different datasets were obtained in: 1) inter-laboratory studies,  
546 2) different sample preparation strategies, 3) different LC gradients and 4) different MS fragmentation  
547 methods. To assess the robustness of the Tc-STAMS2 platform, another batch of *T. cruzi* strains were  
548 processed and acquired using similar chromatographic and MS conditions as described in Figure 1 in

549 an inter-laboratory study perspective. Indeed, the mass spectra library was built with data acquired  
550 in the PR group in Odense, Denmark and the blind samples were acquired in the CEFAP mass  
551 spectrometry facility in São Paulo, Brazil. Although the instrument type and conditions were similar, a  
552 different chromatographic profile was obtained, Supplementary Figure S4. However, the biological  
553 duplicate unknown samples A1 and A2 from DTU-III matched correctly to DTU-III, **Figure 3**. The  
554 unknown sample B from the DTU-I also showed higher similarity scores with the MS/MS spectra  
555 library from DTU-I. The sample B was evaluated on different parameters: 1) sample preparation  
556 conditions such as acid and basic peptide desalting, 2) different chromatographic gradient such as 20  
557 min, 70 min, 130 min, 3) different fragmentation methods, such as CID or HCD with a chromatographic  
558 gradient of 70 min and 4) different sample amount injected into the LC column. Even considering all  
559 these technical sources of variation, the similarity scores continued to match correctly to DTU-I  
560 showing the robustness of the Tc-STAMS2 towards different experimental conditions.

561 In addition, another MS/MS spectral library search software platform, DiagnoProt, was used instead  
562 of SpectraST (Silva *et al.*, 2017b). DiagnoProt was able to differentiate the different DTUs and to  
563 associate the *CL14* strain with the DTU -VI group as shown for the SpectraST software, **Table 4**. Due  
564 to that, two different spectral library search software could be implemented in the Tc-STAMS2  
565 pipeline and used to identify *T. cruzi* strains, **Table 2 and 4**.

566

#### 567 **Clustering Analysis Using Peptide and Protein Identification**

568 Database search was also performed to evaluate the similarity among the DTUs using peptide and  
569 protein identification results. Three database search platforms were used (MaxQuant, TPP and  
570 Proteome Discoverer). From four replicates of each DTU more than 7000 peptides and 4000 proteins  
571 were identified (**Supplementary Figure S5, Supplementary Table S1, S2 and S3**). DTU-I and DTU-VI had  
572 the highest number of identifications due to the protein database used for the search. Indeed, Sylvio  
573 (DTU-I) and CL Brener (DTU-VI) are the two *T. cruzi* strains whose genome has been sequenced and  
574 their proteome annotated and deposited in the Uniprot database. Interestingly, only 30% of the  
575 MS/MS spectra were assigned, leaving behind a wealth of information for *T. cruzi* strain discrimination  
576 (**Supplementary Table S4**).

577 Analysis of variance (ANOVA  $p < 0.05$  followed by Benjamin-Hochberg FDR correction) was  
578 applied for the log<sub>2</sub>-transformed protein or peptide intensities previously identified using MaxQuant.  
579 A total of 1096 proteins and 6130 peptides showed significant difference in abundance among the six  
580 DTUs (**Supplementary Table S5**). The differentially expressed peptides and proteins were subjected to  
581 clustering analysis and visualized as heat-maps, **Supplementary Figure 6A and B**. *CL14* clustered

582 together with CLBrenner (DTU-VI), indicating high similarity in the protein and peptide expression  
583 profile.

584 In addition, principal component analysis (PCA), which was applied in the differential  
585 expressed proteins, were able to discriminate the six DTUs showing a DTU-specific quantitative  
586 proteome repertoire (**Fig 4**). Interestingly, *CL14* and CLBrenner, strains belonging to the DTU-VI, were  
587 also found close to each other, confirming the Euclidean clustering result obtained previously.

588 Additionally, to check the proteomic data reproducibility, a correlation analysis (R squared)  
589 between biological replicates from each DTU or between the different DTUs was performed using  
590 the log<sub>2</sub>-normalized intensities. As shown in **Supplementary Figure S7**, high correlation values were  
591 observed among replicates (R squared>0.9). Interestingly, when we compared different DTUs  
592 (**Supplementary Figure S8**), the R squared dropped to 0.6-0.7, but high correlation was observed  
593 between *CL14* and DTU-VI (R squared = 0.83), confirming once more the similarity between these  
594 two DTUs (**Supplementary Fig S8**).

595

## 596 DISCUSSION

597 Many studies have employed MS-based techniques to identify organisms, such as bacteria  
598 (Patterson e Aebersold, 2003). Önder et, al. pioneered the use of MS/MS spectral libraries to precisely  
599 identify which animal the tick *Ixodes scapularis* was fed even if the feeding occurred months earlier  
600 (Onder *et al.*, 2013). Other strategies combined genetic information in conjunction with MS-based  
601 peptide identification for the correct assignment of microorganisms (Jabbour *et al.*, 2010)  
602 (Dworzanski *et al.*, 2004; Dworzanski *et al.*, 2006). More recently, Shao et al. demonstrated that it is  
603 possible to identify *E. coli* strains with only the MS/MS fragmentation spectra, with no need for  
604 peptide identification (Shao et al. 2015).

605 In the present study, we describe a genome-free, MS/MS spectral-matching methodology  
606 designed to identify different *T. cruzi* DTUs, named Tc-STAMS2. Firstly, a peptide MS/MS spectral  
607 library using three replicates from each of the six DTUs was built. The fourth replicate was used to  
608 test the ability of the method to differentiate each DTUs using the reference library. As shown in **Table**  
609 **2**, this approach was able to differentiate DTUs using the SpectraST software. A unique assignment to  
610 the correct DTU was achieved. The method was tested using a dataset of peptide MS/MS spectra  
611 obtained from different growth conditions of *T.cruzi*.

612 The next step was to test the MS/MS spectral library against phylogenetically related species  
613 such as *T.vivax* and distant organisms such as *E.coli*, *mus musculus* and *homo sapiens* (**Fig. 2**). For such



614 a test, another spectral library database was built with the same DTUs and MS/MS spectra from the  
615 *CL14* strain and *T. vivax*. Interestingly, it is clear that even with thousands of fragmentation spectra,  
616 the scores obtained when comparing samples of human organism or *E.coli* with the library were close  
617 to zero, demonstrating the specificity of the method in identifying only samples of the species/strain,  
618 whose MS/MS spectrum is present in the library.

619 We also showed that the spectral matching method is robust even with inter and intra-  
620 laboratory source of variations using similar MS instruments, but from different laboratory and even  
621 performing changes in the sample preparation, chromatography and fragmentation method, it was  
622 still possible to correctly identify samples from DTU-III and DTU-I, as shown in **Fig. 3**.

623 As expected, we also observed that the similarity score is dependent on the number of MS/MS  
624 spectra acquired. The longer the gradient time used, the greater the separation capacity of peptides  
625 prior to MS/MS analysis and the larger the number of acquired spectra, resulting in higher scores for  
626 the same sample when compared to the library. Using different fragmentation methods we also  
627 observed that CID provides scores higher than HCD when using 70 min gradient time. Although HCD  
628 provides high resolution MS/MS spectra CID provides faster MS/MS sequencing, thus generating  
629 more spectra that can be match with the spectral library.

630 The current methods used to identify DTUs are difficult to implement requiring very well  
631 trained personnel. Other methods require the knowledge of the genome to identify DTU and many  
632 of the methods cannot differentiate strains with similar genome. Our platform is designed to  
633 complement the already developed methodologies and to assist in identification DTUs or any other  
634 parasite. This platform is easy to implement, fast and only requires the MS/MS spectra for  
635 identification.

636 In addition, fragmentation spectra were also subjected to database search analysis for peptide  
637 and protein identification using the MaxQuant software, TPP and Proteome Discoverer. In general,  
638 the number of identifications of proteins and peptides were reproducible and consistent among  
639 different database search software. The larger number of proteins and peptides were observed for  
640 DTU-I and VI. Although the number of annotated proteins in UniProt database is greater for DTUs I  
641 and VI, the number of proteins identified to DTUs V, III and VI, was not significantly smaller. Due to  
642 that, a MS-based proteomic approach can be used to quantitatively compare the protein expression  
643 of different DTUs, even with differences in genome annotation among them and use these  
644 information's to identify DTU-specific pathways correlated with the strain phenotype.

645 Moreover, it was possible to cluster together DTU-VI and *CL14* using the differential expressed  
646 proteins or peptides given by ANOVA test, Supplementary **Fig S6**. This result validates which is already

647 known in the literature, where the *CL14* belong to DTU-VI (Zingales *et al.*, 2009b) and also validates  
648 the results obtained with the spectral matching, where the similarity score between *CL14* and DTU-  
649 VI was higher compared to the other DTUs. Moreover, PCA analysis was also able to discriminate the  
650 six DTUs and determine the similarity between the DTUVI and *CL14*, as seen by the close proximity  
651 between these two groups.

652 Clustering analysis using protein identification in different DTUs was also performed by Telleria  
653 *et al.*, (Telleria *et al.*, 2010) however only 261 (experiment 1) and 172 (experiment 2) 2DE protein  
654 spots were considered for this analysis. In our study, the clustering analysis was performed using 1096  
655 proteins differentially expressed among DTUs, increasing the number of features to build a robust *T.*  
656 *cruzi* clustering (Supplementary Fig S6).

657 The Tc-STAMS2 strategy presented here is robust, accurate, easy to perform and completely  
658 automated. Dworzanski, J. P. (Dworzanski *et al.*, 2006) and Shao, W. *et al.* (Shao *et al.*, 2015) used a  
659 similar methodology for identification of bacteria, however this is the first time that spectral matching  
660 is applied to discriminate different *T. cruzi* DTUs.

661

## 662 CONCLUSION

663 The transmission of Chagas disease through blood transfusion and organ transplant, besides  
664 the triatomine vector, poses several public health challenges. Due to that, novel methods identify and  
665 characterize DTUs can offer opportunities to understand the DTUs diversity, link with their phenotype  
666 and provide another tool for molecular epidemiology. Recently, MALDI-based strategy was developed  
667 for the direct identification of trypanosomatids based on the MS profile, named DIT-MALDI TOF (Avila  
668 *et al.*, 2016). However, this method does not have the resolution to discriminate DTUs using the MS  
669 profile. In this study we present a *T. cruzi* strain typing based on MS/MS spectral matching. The Tc-  
670 STAMS2 method is robust, sensitive and powerful and it is based on the identification of peptides  
671 from their MS/MS spectra. The method could be used to complement other already established  
672 methods. The proposed method can help in the research of epidemiology to identify *T. cruzi* strains  
673 only with the use of fragmentation spectra without the need for genomic data, which is very scarce  
674 for this parasite. Since MS/MS spectra are available to the research community further data mining is  
675 possible in order to improve our understanding on the biology of the different *T. cruzi* strains. The  
676 methodology shown in this study will provide a complementary tool to the current nucleic-based  
677 testing and have the possibility to be extended to other parasitic diseases.

678

## 679 Acknowledgments

680 Marta Campaner is acknowledged for providing genotyped *T.cruzi* strains. Prof. Sirlei Daffre, Joao Alves  
 681 and Claudio R. Marinho are acknowledged for the fruitful discussion. The Biomass facility at CEFAP-  
 682 USP is acknowledged for acquiring some of the LC-MS/MS runs. The Lundbeck foundation is  
 683 acknowledged for the grant sponsoring the Orbitrap Velos Pro. CNPq (GP 441878/2014-8) and FAPESP  
 684 (GP: 2014/06863-3), Rebeca Kawahara is supported by FAPESP (2015/02866-0).

685 **Author Contributions:**

686 **Conceptualization:** GP

687 **Data curation:** GSDO, GP, RK, LRF, MRL

688 **Cell line samples:** MMSGT, CCA

689 **Data Analysis:** GSDO, GP

690 **Methodology:** GSDO, GP

691 **Supervision:** GP

692 **Writing – original draft:** GSDO, RK and GP

693 **Writing – review & editing:** GSDO, RK, LRF, CCA, MMRT, MRL and GP

694 **Tables:**

695 **Table 1. Cultured stocks representing the known *T. cruzi* lineages (DTUs) used to build the MS/MS**  
 696 **library and to validate the method.**

Trypanosomatid	DTUs	Strains	Form	Culture médium
<i>T. cruzi</i>	DTU-I	<i>Sylvio X10/1</i>	Epimastigote	LIT- Liver Infusion- Tryptose supplemented with 10% fetal bovine serum at pH 7,2; 28 °C.
	DTU-II	Y		
	DTU-III	M6241 cl6		
	DTU-IV	CAN III cl1		
	DTU-V	92:80 cl2		
	DTU-VI	CL Brener		
	DTU-VI	Clone 14		

697 Adapted by Camargo, E. (Camargo, E., 1964).

698

699 **Table 2. *T. cruzi* strain identification based on spectral similarity searches. The unique dot product**  
 700 **SDSS score is reported along with the number of MS/MS spectra matches.**

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
DTU-I	0.861 (2858/20232 )	0.005 (11/13450)	0.008 (24/13419)	0.008 (26/14135)	0.016 (65/13470)	0.013 (33/14469)
DTU-II	0.010 (23/14212)	0.754 (1169/20004 )	0.032 (72/14820)	0.023 (62/15126)	0.007 (22/14198)	0.020 (35/16038)
DTU-III	0.006 (18/13177)	0.020 (45/13838)	0.817 (2086/19587 )	0.013 (42/13135)	0.017 (42/13689)	0.015 (35/14200)
DTU-IV	0.011 (38/14307)	0.013 (28/14213)	0.010 (32/13412)	0.849 (2648/20199 )	0.004 (16/12889)	0.019 (48/14722)
DTU-V	0.008 (35/12560)	0.009 (20/12405)	0.013 (37/12771)	0.004 (16/11687)	0.863 (3086/19777 )	0.011 (28/14215)
DTU-VI	0.016 (51/14230)	0.017 (27/14880)	0.013 (32/14101)	0.017 (51/14047)	0.022 (57/14849)	0.800 (1463/20238 )

701 Table made with spectral counting (unique/total), and dot product unique square root (score) of the  
702 result of SpectraST.

703

704

705 Table 3. MS/MS spectral matching comparison using different growth phases of *Sylvio X10/1* (DTU-I).

Samples	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI	CL14 (DTU-VI)
A (St_1)	0.283	0.032	0.068	0.017	0.075	0.032	0.034
A (St_2)	0.288	0.035	0.067	0.015	0.068	0.029	0.030
B (Ex_1)	0.296	0.036	0.057	0.026	0.104	0.058	0.046
B (Ex_2)	0.301	0.041	0.056	0.027	0.099	0.053	0.051

706 Table made with spectral counting (unique/total), and score of the result of SpectraST. St\_1 -  
707 Stationary phase; St\_2- Biological replicate stationary phase; Ex- Exponential phase; Ex\_2- Biological  
708 replicate exponential phase.

709

710

711 Table 4: Spectral matching of different *T.cruzi* strains using the DiagnoProt software

DTUs	DTU-I	DTU-II	DTU-III	DTU-IV	DTU-V	DTU-VI
DTU-I	0,174	0,078	0,082	0,089	0,079	0,094
DTU-II	0,092	0,172	0,093	0,105	0,088	0,104
DTU-III	0,081	0,075	0,162	0,079	0,088	0,092
DTU-IV	0,095	0,084	0,082	0,155	0,079	0,089
DTU-V	0,070	0,065	0,077	0,063	0,155	0,095
DTU-VI	0,088	0,084	0,089	0,077	0,100	0,162
<b>CL14</b>	0,215	0,237	0,246	0,214	0,262	0,360

712 Table 4- Comparison of the spectral library against the 6 DTUs and *CL14* strain. The database was  
 713 constructed with six DTUs (DTU-I – DTU-VI) after the library was constructed and it was used for  
 714 comparison of the DTUs and *CL14*, which is not included in the database.

715

716 **Figures:**

717 **Figure 1: Tc-STAMS2 workflow.** (A) Proteins were extracted from *T. cruzi* cells, digested and subjected  
 718 to nLC-MS/MS analysis. MS/MS spectra were clustered and merged to generate a reference mass  
 719 spectral library. (B) The same steps described in panel A were used for the unknown samples. (C) The  
 720 generated spectra were compared with the reference library in order to assign every unknown  
 721 sample to a particular *T.cruzi* strain. (D) The data obtained by MSMS mass spectrometry were  
 722 searched against the *T.cruzi* protein database.

723

724 **Figure 2:** Tc-STAMS2 approach tested against : 1) the *CL14* *T.cruzi* strain, 2) *T. vivax* dataset and 3) LC-  
 725 MS/MS datasets from *E.coli* and human placental tissues.

726

727 **Fig 3. Tc-STAMS2 was tested for its robustness towards technical and experimental variations.** Two  
 728 unknown *T.cruzi* strains (A and B) were processed as described in figure 1 and acquired in the CEFAP  
 729 mass spectrometry facility in Sao Paulo, Brazil. The spectral library was built with data acquired in  
 730 the PR group, Odense, Denmark using a similar LC-MS/MS setup. A1 and A2 are a biological  
 731 duplicate of *T.cruzi* DTU-III. B is the *T.cruzi* DTU-I. B/acid: Peptides derived from sample B were  
 732 purified using acidic conditions (0.1% TFA). B/basic: Peptides derived from sample B were purified  
 733 using basic conditions (0.1% ammonia water). CID- Collision-Induced Dissociation and HCD- Higher-  
 734 energy collisional dissociation. High and Low indicate the amount of sample injected into the LC-MS  
 735 system, 1 and 0.5 ug, respectively.

736

737 **Fig. 4 -** Principal component analysis (PCA) of the proteins of the six DTU plus *CL14*. DTU I- Red, DTU  
 738 II- Blue, DTU III- Yellow, DTU IV- Orange, DTU V – Black, DTU VI – Green and *CL14*- Grey.

739

740 **Supplementary figures:**

741 **Supplementary Figure S1:** LC-MS chromatographic profile of peptides belonging to the four  
 742 replicates from DTU I. The Pearson correlation score is reported on the right side of the  
 743 chromatogram.

744

745 **Supplementary Figure S2:** LC-MS chromatographic profile of peptides belonging to the six *T.cruzi*  
 746 strains.

747

748 **Supplementary Figure S3:** LC-MS chromatographic profile of peptides belonging to the CLBrenner and  
 749 *CL14* *T.cruzi* strains.

750

751 **Supplementary Figure S4:** LC-MS chromatographic profile of peptides belonging to the DTU-I

752 acquired in the PR Group, Odense, Denmark and in the CEFAP mass spectrometry facility at USP, Sao  
753 Paulo, Brazil.  
754

755 **Supplementary Fig S5.** A. Peptide identifications using MaxQuant, TPP and Proteome Discoverer™  
756 softwares. B. Protein identifications using MaxQuant, TPP and Proteome Discoverer™ softwares.

757 **Supplementary Figure S6:** Heat map of differentially regulated A) peptides and B) proteins.  
758

759 **Supplementary Figure S7-** Multi Scatter Plot of the six DTUs with four technical replicates each using  
760 peptides. A – DTU I, B – DTU II, C – DTU III, D – DTU IV, E – DTU V and F – DTU VI.  
761

762 **Supplementary Figure S8-** Multi Scatter Plot constructed using the peptides with all DTUs + *CL14*,  
763 using only a technical replica for each DTU.  
764

765

766 **Supplementary tables:**

767 **Supplementary table S1, S2 and S3:** number of protein IDs using the MaxQuant, TPP and Proteome  
768 Discoverer platforms.

769 **Supplementary table S4:** Number of PSMs and unannotated MS/MS spectra.

770 **Supplementary table S5:** Regulated peptides and proteins among the six DTUs.  
771

771

772 **References**

773

774 A PODLIPAEV, S.; FROLOV, A. [**The phylogeny of the Trypanosomatidae: the molecular and**  
775 **morphological approaches**]. 2000. 169-82.  
776

776

777 AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. **Nature**, v. 422, n. 6928, p.  
778 198-207, Mar 13 2003. ISSN 0028-0836 (Print)

779 0028-0836 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12634793> >.  
780

780

781 AGABIAN, N. Trans splicing of nuclear pre-mRNAs. **Cell**, v. 61, n. 7, p. 1157-60, Jun 29 1990. ISSN  
782 0092-8674 (Print)

783 0092-8674 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/2142018> >.  
784

784

785 ALTELAAR, A. F.; MUNOZ, J.; HECK, A. J. Next-generation proteomics: towards an integrative  
786 view of proteome dynamics. **Nat Rev Genet**, v. 14, n. 1, p. 35-48, Jan 2013. ISSN 1471-0064  
787 (Electronic)

788 1471-0056 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23207911> >.  
789

789

790 ANDERSON, N. L. et al. Global approaches to quantitative analysis of gene-expression patterns  
791 observed by use of two-dimensional gel electrophoresis. **Clin Chem**, v. 30, n. 12 Pt 1, p. 2031-6, Dec  
792 1984. ISSN 0009-9147 (Print)

793 0009-9147 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6499176> >.  
794

794

795 ANDRADE, D. V.; GOLLOB, K. J.; DUTRA, W. O. Acute chagas disease: new global challenges  
796 for an old neglected disease. **PLoS Negl Trop Dis**, v. 8, n. 7, p. e3010, 2014. ISSN 1935-2735  
797 (Electronic)

798 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25077613> >.  
799

799

- 800 ANDRADE, S. G. Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano. In:  
801 (Ed.). **Revista Patologia Tropical**, v.3, 1974. p.65-121.  
802
- 803 ANDREWS, N. W. Living dangerously: how *Trypanosoma cruzi* uses lysosomes to get inside host  
804 cells, and then escapes into the cytoplasm. **Biol Res**, v. 26, n. 1-2, p. 65-7, 1993. ISSN 0716-9760  
805 (Print)  
806 0716-9760 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7670547> >.  
807
- 808 ANTAS, P. R. et al. Early, intermediate, and late acute stages in Chagas' disease: a study combining  
809 anti-galactose IgG, specific serodiagnosis, and polymerase chain reaction analysis. **Am J Trop Med**  
810 **Hyg**, v. 61, n. 2, p. 308-14, Aug 1999. ISSN 0002-9637 (Print)  
811 0002-9637 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10463685> >.  
812
- 813 ARAUJO, P. R.; TEIXEIRA, S. M. Regulatory elements involved in the post-transcriptional control  
814 of stage-specific gene expression in *Trypanosoma cruzi*: a review. **Mem Inst Oswaldo Cruz**, v. 106,  
815 n. 3, p. 257-66, May 2011. ISSN 1678-8060 (Electronic)  
816 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21655811> >.  
817
- 818 ATAYDE, V. D. et al. Molecular basis of non-virulence of *Trypanosoma cruzi* clone CL-14. **Int J**  
819 **Parasitol**, v. 34, n. 7, p. 851-60, Jun 2004. ISSN 0020-7519 (Print)  
820 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15157768> >.  
821
- 822 ATWOOD, J. A., 3RD et al. The *Trypanosoma cruzi* proteome. **Science**, v. 309, n. 5733, p. 473-6,  
823 Jul 15 2005. ISSN 1095-9203 (Electronic)  
824 0036-8075 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16020736> >.  
825
- 826 ATWOOD, J. A. et al. The *Trypanosoma cruzi* Proteome. **Science**, v. 309, n. 5733, p. 473-476, 2005-  
827 07-15 00:00:00 2005. Disponível em: < <http://science.sciencemag.org/sci/309/5733/473.full.pdf> >.  
828
- 829 AVILA, C. C.; ALMEIDA, F. G.; PALMISANO, G. Direct identification of trypanosomatids by  
830 matrix-assisted laser desorption ionization-time of flight mass spectrometry (DIT MALDI-TOF MS).  
831 **J Mass Spectrom**, v. 51, n. 8, p. 549-57, Aug 2016. ISSN 1096-9888 (Electronic)  
832 1076-5174 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27659938> >.  
833
- 834 BAKER, J. R. et al. Biochemical characterization of some species of *Trypanosoma* (*Schizotrypanum*)  
835 from bats (Microchiroptera). **Am J Trop Med Hyg**, v. 27, n. 3, p. 483-91, May 1978. ISSN 0002-  
836 9637 (Print)  
837 0002-9637 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/354417> >.  
838
- 839 BARBOSA, E. B. et al. Proteômica: metodologias e aplicações no estudo de doenças humanas.  
840 **Revista da Associação Médica Brasileira**, v. 58, n. 3, p. 366-375, 2012/05/01/ 2012. ISSN 0104-  
841 4230. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0104423012705237> >.  
842
- 843 BARNABE, C.; BRISSE, S.; TIBAYRENC, M. Phylogenetic diversity of bat trypanosomes of  
844 subgenus *Schizotrypanum* based on multilocus enzyme electrophoresis, random amplified  
845 polymorphic DNA, and cytochrome b nucleotide sequence analyses. **Infect Genet Evol**, v. 2, n. 3, p.  
846 201-8, Feb 2003. ISSN 1567-1348 (Print)  
847 1567-1348 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12797982> >.  
848
- 849 BARRETO, M. L. et al. Successes and failures in the control of infectious diseases in Brazil: social  
850 and environmental context, policies, interventions, and research needs. **Lancet**, v. 377, n. 9780, p.  
851 1877-89, May 28 2011. ISSN 1474-547X (Electronic)

- 852 0140-6736 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21561657> >.  
853
- 854 BARTHOLOMEU, D. C. et al. Unveiling the intracellular survival gene kit of trypanosomatid  
855 parasites. **PLoS Pathog**, v. 10, n. 12, p. e1004399, Dec 2014. ISSN 1553-7374 (Electronic)  
856 1553-7366 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25474314> >.  
857
- 858 BATISTA, J. A. et al. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes.  
859 **Mol Biochem Parasitol**, v. 67, n. 2, p. 301-12, Oct 1994. ISSN 0166-6851 (Print)  
860 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7870134> >.  
861
- 862 BENSIMON, A.; HECK, A. J.; AEBERSOLD, R. Mass spectrometry-based proteomics and network  
863 biology. **Annu Rev Biochem**, v. 81, p. 379-405, 2012. ISSN 1545-4509 (Electronic)  
864 0066-4154 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22439968> >.  
865
- 866 BERN, C. et al. Evaluation and treatment of chagas disease in the United States: a systematic review.  
867 **JAMA**, v. 298, n. 18, p. 2171-81, Nov 14 2007. ISSN 1538-3598 (Electronic)  
868 0098-7484 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18000201> >.  
869
- 870 BONALDO, M. C. et al. Cell-substrate adhesion during *Trypanosoma cruzi* differentiation. **J Cell**  
871 **Biol**, v. 106, n. 4, p. 1349-58, Apr 1988. ISSN 0021-9525 (Print)  
872 0021-9525 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/3283152> >.  
873
- 874 BRENER, Z. *Trypanosoma cruzi*: morfologia e ciclo evolutivo. In: JOÃO CARLOS PINTO DIAS,  
875 J. R. C. (Ed.). **Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico**  
876 **geral**. Scielo Books: FIOCRUZ, 1997. cap. 24-31, p.486. ISBN ISBN 85-85676-31-0.  
877
- 878 BRENER, Z.; CHIARI, E. [Morphological Variations Observed in Different Strains of *Trypanosoma*  
879 *Cruzi*]. **Rev Inst Med Trop Sao Paulo**, v. 5, p. 220-4, Sep-Oct 1963a. ISSN 0036-4665 (Print)  
880 0036-4665 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/14110094> >.  
881
- 882 \_\_\_\_\_. Variações Morfológicas Observadas em Diferentes Amostras de *Trypanosoma cruzi*. **Rev**  
883 **Inst Med Trop Sao Paulo**, 1963b.  
884
- 885 BRISSE, S. et al. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener  
886 reference strain by multilocus enzyme electrophoresis and multiprimer random amplified  
887 polymorphic DNA fingerprinting. **Mol Biochem Parasitol**, v. 92, n. 2, p. 253-63, May 01 1998. ISSN  
888 0166-6851 (Print)  
889 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9657330> >.  
890
- 891 BRISSE, S.; VERHOEF, J.; TIBAYRENC, M. Characterisation of large and small subunit rRNA and  
892 mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. **International**  
893 **Journal for Parasitology**, v. 31, n. 11, p. 1218-26, Sep 2001a. ISSN 0020-7519 (Print)  
894 0020-7519 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11513891> >.  
895
- 896 \_\_\_\_\_. Characterisation of large and small subunit rRNA and mini-exon genes further supports the  
897 distinction of six *Trypanosoma cruzi* lineages. **Int J Parasitol**, v. 31, n. 11, p. 1218-26, Sep 2001b.  
898 ISSN 0020-7519 (Print)  
899 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11513891> >.  
900
- 901 BRUCE ALBERTS, A. J., JULIAN LEWIS, MARTIN RAFF, KEITH ROBERTS, AND PETER  
902 WALTE. **Molecular Biology of the Cell**. Garland Science, 2002. ISBN 0-8153-3218-1 0-8153-  
903 4072-9. Disponível em: < <https://www.ncbi.nlm.nih.gov/books/NBK21054/> >.



- 904  
905 BRUNORO, G. V. F. et al. Reevaluating the *Trypanosoma cruzi* proteomic map: The shotgun  
906 description of bloodstream trypomastigotes. **Journal of Proteomics**, v. 115, p. 58-65, 2/6/ 2015.  
907 ISSN 1874-3919. Disponível em: <  
908 <http://www.sciencedirect.com/science/article/pii/S1874391914005569>>.  
909
- 910 BYTHELL, B. J. et al. Proton-driven amide bond-cleavage pathways of gas-phase peptide ions  
911 lacking mobile protons. **J Am Chem Soc**, v. 131, n. 39, p. 14057-65, Oct 07 2009. ISSN 1520-5126  
912 (Electronic)  
913 0002-7863 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19746933> >.  
914
- 915 CAMARGO, E. Growth and differentiation in *Trypanosoma cruzi*. i. origin of metacyclic  
916 trypanosomes in liquid media (1964). **Revista do Instituto de Medicina Tropical de São Paulo**, v.  
917 12, p. 93-100, 1964.  
918
- 919 CAMARGO, E. P. Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of Metacyclic  
920 Trypanosomes in Liquid Media. **Rev Inst Med Trop Sao Paulo**, v. 6, p. 93-100, May-Jun 1964. ISSN  
921 0036-4665 (Print)  
922 0036-4665 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/14177814> >.  
923
- 924 CANTÚ, M. D. et al. Seqüenciamento de peptídeos usando espectrometria de massas: um guia  
925 prático. **Química Nova**, v. 31, p. 669-675, 2008. ISSN 0100-4042. Disponível em: <  
926 [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40422008000300034&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422008000300034&nrm=iso) >.  
927
- 928 CARLIER, Y. et al. Congenital Chagas disease: recommendations for diagnosis, treatment and  
929 control of newborns, siblings and pregnant women. **PLoS Negl Trop Dis**, v. 5, n. 10, p. e1250, Oct  
930 2011. ISSN 1935-2735 (Electronic)  
931 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22039554> >.  
932
- 933 CARR, S. A.; HUDDLESTON, M. J.; BEAN, M. F. Selective identification and differentiation of N-  
934 and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry.  
935 **Protein Sci**, v. 2, n. 2, p. 183-96, Feb 1993. ISSN 0961-8368 (Print)  
936 0961-8368 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7680267> >.  
937
- 938 CASS, Q.; CASSIANO, N. **Cromatografia Líquida: Novas Tendências e Aplicações**. Elsevier  
939 Academic, 2015. ISBN ISBN13:9788535275971; ISBN10:8535275975  
940
- 941 CAVALIER-SMITH, T. Eukaryote kingdoms: Seven or nine? **Biosystems**, v. 14, n. 3, p. 461-481,  
942 1981/01/01/ 1981. ISSN 0303-2647. Disponível em: <  
943 <http://www.sciencedirect.com/science/article/pii/0303264781900502> >.  
944
- 945 CDC. Parasites - American Trypanosomiasis (also known as Chagas Disease). **Center for Disease**  
946 **Control and Prevention**, Center for Disease Control and Prevention, 2017. Disponível em: <  
947 <https://www.cdc.gov/parasites/chagas/biology.html> >.  
948
- 949 CHAGAS, C. Nova tripanozomíaze humana: estudos sobre a morfologia e o ciclo evolutivo do  
950 *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem.  
951 **Memórias do Instituto Oswaldo Cruz**, v. 1, p. 159-218, 1909. ISSN 0074-0276. Disponível em: <  
952 [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0074-02761909000200008&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02761909000200008&nrm=iso) >.  
953

- 954 CHANDRAMOULI, K.; QIAN, P. Y. Proteomics: challenges, techniques and possibilities to  
955 overcome biological sample complexity. **Hum Genomics Proteomics**, v. 2009, Dec 08 2009. ISSN  
956 1757-4242 (Electronic)  
957 1757-4242 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20948568> >.  
958
- 959 CORTHALS, G. L. et al. The dynamic range of protein expression: a challenge for proteomic  
960 research. **Electrophoresis**, v. 21, n. 6, p. 1104-15, Apr 2000. ISSN 0173-0835 (Print)  
961 0173-0835 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10786884> >.  
962
- 963 COSENTINO, R. O.; AGUERO, F. A simple strain typing assay for *Trypanosoma cruzi*:  
964 discrimination of major evolutionary lineages from a single amplification product. **PLoS Neglected**  
965 **Tropical Diseases**, v. 6, n. 7, p. e1777, 2012a.  
966
- 967 \_\_\_\_\_. A simple strain typing assay for *Trypanosoma cruzi*: discrimination of major evolutionary  
968 lineages from a single amplification product. **PLoS Negl Trop Dis**, v. 6, n. 7, p. e1777, 2012b. ISSN  
969 1935-2735 (Electronic)  
970 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22860154> >.  
971
- 972 COURA, J. R.; DIAS, J. C. Epidemiology, control and surveillance of Chagas disease: 100 years after  
973 its discovery. **Mem Inst Oswaldo Cruz**, v. 104 Suppl 1, p. 31-40, Jul 2009. ISSN 1678-8060  
974 (Electronic)  
975 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19753455> >.  
976
- 977 COX, J.; MANN, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range  
978 mass accuracies and proteome-wide protein quantification. **Nat Biotechnol**, v. 26, n. 12, p. 1367-72,  
979 Dec 2008. ISSN 1546-1696 (Electronic)  
980 1087-0156 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19029910> >.  
981
- 982 COX, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. **J**  
983 **Proteome Res**, v. 10, n. 4, p. 1794-805, Apr 01 2011. ISSN 1535-3907 (Electronic)  
984 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21254760> >.  
985
- 986 CRAIG, R. et al. Using Annotated Peptide Mass Spectrum Libraries for Protein Identification.  
987 **Journal of Proteome Research**, v. 5, n. 8, p. 1843-1849, 2006/08/01 2006. ISSN 1535-3893.  
988 Disponível em: < <http://dx.doi.org/10.1021/pr0602085> >.  
989
- 990 CRAIG, R.; CORTENS, J. P.; BEAVIS, R. C. Open source system for analyzing, validating, and  
991 storing protein identification data. **J Proteome Res**, v. 3, n. 6, p. 1234-42, Nov-Dec 2004. ISSN 1535-  
992 3893 (Print)  
993 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15595733> >.  
994
- 995 CURA, C. I. et al. Multiplex Real-Time PCR Assay Using TaqMan Probes for the Identification of  
996 *Trypanosoma cruzi* DTUs in Biological and Clinical Samples. **PLoS Negl Trop Dis**, v. 9, n. 5, p.  
997 e0003765, May 2015. ISSN 1935-2735 (Electronic)  
998 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25993316> >.  
999
- 1000 CURA, C. I. et al. Multiplex Real-Time PCR Assay Using TaqMan Probes for the Identification of  
1001 *Trypanosoma cruzi* DTUs in Biological and Clinical Samples. **PLoS Negl Trop Dis**,  
1002 v. 9, n. 5, p. e0003765, 2015. Disponível em: < <http://dx.doi.org/10.1371/journal.pntd.0003765> >.  
1003

- 1004 DA NOBREGA, A. A.; DE ARAUJO, W. N.; VASCONCELOS, A. M. Mortality due to Chagas  
1005 disease in Brazil according to a specific cause. **Am J Trop Med Hyg**, v. 91, n. 3, p. 528-33, Sep 2014.  
1006 ISSN 1476-1645 (Electronic)  
1007 0002-9637 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25002301> >.  
1008
- 1009 DAS, A. Spinning Charged Test-Particles in General Relativity. **Progress of Theoretical Physics**, v.  
1010 23, n. 4, p. 610-615, 1960. ISSN 0033-068X. Disponível em: <  
1011 <http://dx.doi.org/10.1143/PTP.23.610> >.  
1012
- 1013 DASS, C. Mass Analysis and Ion Detection. In: (Ed.). **Fundamentals of Contemporary Mass**  
1014 **Spectrometry**: John Wiley & Sons, Inc., 2006. p.67-117. ISBN 9780470118498.  
1015
- 1016 DE GAUDENZI, J. G. et al. Gene expression regulation in trypanosomatids. **Essays Biochem**, v. 51,  
1017 p. 31-46, 2011. ISSN 1744-1358 (Electronic)  
1018 0071-1365 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22023440> >.  
1019
- 1020 DE GODOY, L. M. et al. Quantitative proteomics of Trypanosoma cruzi during metacyclogenesis.  
1021 **Proteomics**, v. 12, n. 17, p. 2694-703, Aug 2012. ISSN 1615-9853.  
1022
- 1023 DE HOOG, C. L.; MANN, M. Proteomics. **Annu Rev Genomics Hum Genet**, v. 5, p. 267-93, 2004.  
1024 ISSN 1527-8204 (Print)  
1025 1527-8204 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15485350> >.  
1026
- 1027 DESIERE, F. et al. The PeptideAtlas project. **Nucleic Acids Res**, v. 34, n. Database issue, p. D655-  
1028 8, Jan 01 2006. ISSN 1362-4962 (Electronic)  
1029 0305-1048 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16381952> >.  
1030
- 1031 DESIERE, F. et al. Integration with the human genome of peptide sequences obtained by high-  
1032 throughput mass spectrometry. **Genome Biol**, v. 6, n. 1, p. R9, 2005. ISSN 1474-760X (Electronic)  
1033 1474-7596 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15642101> >.  
1034
- 1035 DEUTSCH, E. W. et al. A guided tour of the trans-proteomic pipeline. **Proteomics**, v. 10, n. 6, p.  
1036 1150-9, Mar 2010. ISSN 1615-9861 (Electronic)  
1037 1615-9853 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20101611> >.  
1038
- 1039 DIAS, J.; SILVEIRA, A.; SCHOFIELD, C. The impact of Chagas disease control in Latin America:  
1040 a review. **Memórias do Instituto Oswaldo Cruz**, v. 97, p. 603-612, 2002. ISSN 0074-0276.  
1041 Disponível em: < [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0074-  
1042 02762002000500002&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762002000500002&nrm=iso) >.  
1043
- 1044 DIAS, J. C.; COURA, J. **Clínica e Terapêutica da Doença de Chagas: uma abordagem prática**  
1045 **para o clínico geral**. FIOCRUZ: FIOCRUZ, 1997. ISBN 85-85676- 31-0. Disponível em: <  
1046 [https://portal.fiocruz.br/pt-br/content/clinica-e-terapeutica-da-doenca-de-chagas-uma-abordagem-  
1047 pratica-para-o-clinico-geral](https://portal.fiocruz.br/pt-br/content/clinica-e-terapeutica-da-doenca-de-chagas-uma-abordagem-pratica-para-o-clinico-geral) >.  
1048
- 1049 DIAS, J. C. et al. [Brazilian Consensus on Chagas Disease, 2015]. **Epidemiol Serv Saude**, v. 25, n.  
1050 spe, p. 7-86, Jun 2016. ISSN 2237-9622 (Electronic)  
1051 1679-4974 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27869914> >.  
1052
- 1053 DIAS, J. C. P.; COURA, J. R.; YASUDA, M. A. S. The present situation, challenges, and perspectives  
1054 regarding the production and utilization of effective drugs against human Chagas disease. **Revista da**  
1055 **Sociedade Brasileira de Medicina Tropical**, v. 47, p. 123-125, 2014. ISSN 0037-8682. Disponível

- 1056 em: < [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0037-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-)  
1057 [86822014000100123&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86822014000100123&nrm=iso) >.  
1058
- 1059 DIAZ, M. L.; TORRES, R.; GONZALEZ, C. I. [Differential protein expression in developmental  
1060 stages of *Trypanosoma cruzi* I isolated from a patient with chronic chagasic cardiomyopathy].  
1061 **Biomedica**, v. 31, n. 4, p. 503-13, Oct-Dec 2011. ISSN 0120-4157 (Print)  
1062 0120-4157 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22674361> >.  
1063
- 1064 DOCAMPO, R. Sensitivity of parasites to free radical damage by antiparasitic drugs. **Chem Biol**  
1065 **Interact**, v. 73, n. 1, p. 1-27, 1990. ISSN 0009-2797 (Print)  
1066 0009-2797 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/2406032> >.  
1067
- 1068 DOMON, B.; AEBERSOLD, R. Mass spectrometry and protein analysis. **Science**, v. 312, n. 5771, p.  
1069 212-7, Apr 14 2006. ISSN 1095-9203 (Electronic)  
1070 0036-8075 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16614208> >.  
1071
- 1072 DOS SANTOS JUNIOR ADE, C. et al. Unveiling the *Trypanosoma cruzi* Nuclear Proteome. **PLoS**  
1073 **One**, v. 10, n. 9, p. e0138667, 2015. ISSN 1932-6203 (Electronic)  
1074 1932-6203 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26383644> >.  
1075
- 1076 DWORZANSKI, J. P. et al. Mass spectrometry-based proteomics combined with bioinformatic tools  
1077 for bacterial classification. **J Proteome Res**, v. 5, n. 1, p. 76-87, Jan 2006. ISSN 1535-3893 (Print)  
1078 1535-3893 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16396497> >.  
1079
- 1080 DWORZANSKI, J. P. et al. Identification of bacteria using tandem mass spectrometry combined  
1081 with a proteome database and statistical scoring. **Anal Chem**, v. 76, n. 8, p. 2355-66, Apr 15 2004.  
1082 ISSN 0003-2700 (Print)  
1083 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15080748> >.  
1084
- 1085 EBERT, F. Comparison of isoenzymes of some species of the subgenus *schizotrypanum* from bats by  
1086 isoelectrofocusing. **Tropenmed Parasitol**, v. 34, n. 2, p. 93-7, Jun 1983. ISSN 0303-4208 (Print)  
1087 0303-4208 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6224326> >.  
1088
- 1089 EFRON, B.; ROBERT. **An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs**  
1090 **on Statistics & Applied Probability)**. Chapman and Hall/CRC, 1994. ISBN ISBN-10: 0412042312;  
1091 ISBN-13: 978-0412042317.  
1092
- 1093 EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas  
1094 disease. **Science**, v. 309, n. 5733, p. 409-15, Jul 15 2005. ISSN 1095-9203 (Electronic)  
1095 0036-8075 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16020725> >.  
1096
- 1097 ENG, J. K.; JAHAN, T. A.; HOOPMANN, M. R. Comet: an open-source MS/MS sequence database  
1098 search tool. **Proteomics**, v. 13, n. 1, p. 22-4, Jan 2013. ISSN 1615-9861 (Electronic)  
1099 1615-9853 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23148064> >.  
1100
- 1101 ENG, J. K.; MCCORMACK, A. L.; YATES, J. R. An approach to correlate tandem mass spectral data  
1102 of peptides with amino acid sequences in a protein database. **J Am Soc Mass Spectrom**, v. 5, n. 11,  
1103 p. 976-89, Nov 1994. ISSN 1044-0305 (Print)  
1104 1044-0305 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24226387> >.  
1105
- 1106 FERELLA, M. et al. Proteomics in *Trypanosoma cruzi*--localization of novel proteins to various  
1107 organelles. **Proteomics**, v. 8, n. 13, p. 2735-49, Jul 2008. ISSN 1615-9861 (Electronic)

- 1108 1615-9853 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18546153> >.  
1109
- 1110 FILARDI, L. S.; BRENER, Z. A rapid method for testing in vivo the susceptibility of different strains  
1111 of *Trypanosoma cruzi* to active chemotherapeutic agents. **Mem Inst Oswaldo Cruz**, v. 79, n. 2, p.  
1112 221-5, Apr-Jun 1984. ISSN 0074-0276 (Print)  
1113 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6443014> >.  
1114
- 1115 FRANZEN, O. et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I *Sylvio X10/1* and  
1116 comparison with *T. cruzi* VI CL Brener. **PLoS Negl Trop Dis**, v. 5, n. 3, p. e984, Mar 08 2011. ISSN  
1117 1935-2735 (Electronic)  
1118 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21408126> >.  
1119
- 1120 FRANZEN, O. et al. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages  
1121 with the bat-restricted subspecies *T. cruzi marinkellei*. **BMC Genomics**, v. 13, p. 531, Oct 05 2012.  
1122 ISSN 1471-2164 (Electronic)  
1123 1471-2164 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23035642> >.  
1124
- 1125 FREWEN, B. E. et al. Analysis of peptide MS/MS spectra from large-scale proteomics experiments  
1126 using spectrum libraries. **Anal Chem**, v. 78, n. 16, p. 5678-84, Aug 15 2006. ISSN 0003-2700 (Print)  
1127 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16906711> >.  
1128
- 1129 GARRARD, E. A. et al. Inhibition of Trypanothione Reductase by Substrate Analogues. **Organic**  
1130 **Letters**, v. 2, n. 23, p. 3639-3642, 2000/11/01 2000. ISSN 1523-7060. Disponível em: <  
1131 <http://dx.doi.org/10.1021/ol0065423> >.  
1132
- 1133 GIESE, V. et al. *Trypanosoma cruzi*: a stage-specific calpain-like protein is induced after various  
1134 kinds of stress. **Mem Inst Oswaldo Cruz**, v. 103, n. 6, p. 598-601, Sep 2008. ISSN 1678-8060  
1135 (Electronic)  
1136 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18949332> >.  
1137
- 1138 GLISH, G. L.; VACHET, R. W. The basics of mass spectrometry in the twenty-first century. **Nat Rev**  
1139 **Drug Discov**, v. 2, n. 2, p. 140-50, Feb 2003. ISSN 1474-1776 (Print)  
1140 1474-1776 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12563305> >.  
1141
- 1142 GOMES, M. L. et al. Chagas' disease diagnosis: comparative analysis of parasitologic, molecular,  
1143 and serologic methods. **Am J Trop Med Hyg**, v. 60, n. 2, p. 205-10, Feb 1999. ISSN 0002-9637  
1144 (Print)  
1145 0002-9637 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10072137> >.  
1146
- 1147 GRISARD, E. C. et al. *Trypanosoma cruzi* Clone Dm28c Draft Genome Sequence. **Genome**  
1148 **Announc**, v. 2, n. 1, Jan 30 2014. ISSN 2169-8287 (Print). Disponível em: <  
1149 <https://www.ncbi.nlm.nih.gov/pubmed/24482508> >.  
1150
- 1151 HAJDUK, S.; OCHSENREITER, T. RNA editing in kinetoplastids. **RNA Biol**, v. 7, n. 2, p. 229-36,  
1152 Mar-Apr 2010. ISSN 1555-8584 (Electronic)  
1153 1547-6286 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20220308> >.  
1154
- 1155 HART-SMITH, G.; BLANKSBY, S. J. Mass Analysis. In: (Ed.). **Mass Spectrometry in Polymer**  
1156 **Chemistry**: Wiley-VCH Verlag GmbH & Co. KGaA, 2011. p.5-32. ISBN 9783527641826.  
1157

- 1158 HENRIKSSON, J. et al. Chromosomal localization of seven cloned antigen genes provides evidence  
1159 of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi*. **Mol Biochem**  
1160 **Parasitol**, v. 42, n. 2, p. 213-23, Sep-Oct 1990a. ISSN 0166-6851 (Print)  
1161 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/2270104> >.  
1162
- 1163 \_\_\_\_\_ . Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and  
1164 further demonstration of karyotype variability in *Trypanosoma cruzi*. **Molecular and Biochemical**  
1165 **Parasitology**, v. 42, n. 2, p. 213-23, Sep-Oct 1990b. ISSN 0166-6851 (Print)  
1166 0166-6851 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2270104> >.  
1167
- 1168 HILTON, G. R.; BENESCH, J. L. Two decades of studying non-covalent biomolecular assemblies  
1169 by means of electrospray ionization mass spectrometry. **J R Soc Interface**, v. 9, n. 70, p. 801-16,  
1170 May 07 2012. ISSN 1742-5662 (Electronic)  
1171 1742-5662 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22319100> >.  
1172
- 1173 HOTEZ, P. J.; FUJIWARA, R. T. Brazil's neglected tropical diseases: an overview and a report card.  
1174 **Microbes Infect**, v. 16, n. 8, p. 601-6, Aug 2014. ISSN 1769-714X (Electronic)  
1175 1286-4579 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25088506> >.  
1176
- 1177 HU, Q. et al. The Orbitrap: a new mass spectrometer. **J Mass Spectrom**, v. 40, n. 4, p. 430-43, Apr  
1178 2005. ISSN 1076-5174 (Print)  
1179 1076-5174 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15838939> >.  
1180
- 1181 HUNT, D. F. et al. Protein sequencing by tandem mass spectrometry. **Proc Natl Acad Sci U S A**, v.  
1182 83, n. 17, p. 6233-7, Sep 1986. ISSN 0027-8424 (Print)  
1183 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/3462691> >.  
1184
- 1185 ISKEN, O.; MAQUAT, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal  
1186 mRNA function. **Genes Dev**, v. 21, n. 15, p. 1833-56, Aug 01 2007. ISSN 0890-9369 (Print)  
1187 0890-9369 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17671086> >.  
1188
- 1189 JABBOUR, R. E. et al. Double-blind characterization of non-genome-sequenced bacteria by mass  
1190 spectrometry-based proteomics. **Appl Environ Microbiol**, v. 76, n. 11, p. 3637-44, Jun 2010. ISSN  
1191 1098-5336 (Electronic)  
1192 0099-2240 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20363779> >.  
1193
- 1194 JACKSON, A. P. et al. Global Gene Expression Profiling through the Complete Life Cycle of  
1195 *Trypanosoma vivax*. **PLoS Negl Trop Dis**, v. 9, n. 8, p. e0003975, 2015. ISSN 1935-2735 (Electronic)  
1196 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26266535> >.  
1197
- 1198 JANKEVICIUS, J. V. et al. [Biological cycle of *Phytomonas*]. **Mem Inst Oswaldo Cruz**, v. 83 Suppl  
1199 1, p. 601-10, Nov 1988. ISSN 0074-0276 (Print)  
1200 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/3253512> >.  
1201
- 1202 JOHNSON, J. V. et al. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles  
1203 and quadrupole ion traps. **Analytical Chemistry**, v. 62, n. 20, p. 2162-2172, 1990/10/15 1990. ISSN  
1204 0003-2700. Disponível em: < <http://dx.doi.org/10.1021/ac00219a003> >.  
1205
- 1206 JOHNSON, P. J.; KOOTER, J. M.; BORST, P. Inactivation of transcription by UV irradiation of *T.*  
1207 *brucei* provides evidence for a multicistronic transcription unit including a VSG gene. **Cell**, v. 51, n.  
1208 2, p. 273-81, Oct 23 1987. ISSN 0092-8674 (Print)  
1209 0092-8674 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/3664637> >.

- 1210  
1211 KEBARLE, P. A brief overview of the present status of the mechanisms involved in electrospray mass  
1212 spectrometry. **J Mass Spectrom**, v. 35, n. 7, p. 804-17, Jul 2000. ISSN 1076-5174 (Print)  
1213 1076-5174 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10934434> >.  
1214
- 1215 KELLER, A. et al. A uniform proteomics MS/MS analysis platform utilizing open XML file formats.  
1216 **Mol Syst Biol**, v. 1, p. 2005 0017, 2005. ISSN 1744-4292 (Electronic)  
1217 1744-4292 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16729052> >.  
1218
- 1219 KELLER, A. et al. Empirical statistical model to estimate the accuracy of peptide identifications  
1220 made by MS/MS and database search. **Anal Chem**, v. 74, n. 20, p. 5383-92, Oct 15 2002. ISSN 0003-  
1221 2700 (Print)  
1222 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12403597> >.  
1223
- 1224 KESSNER, D. et al. ProteoWizard: open source software for rapid proteomics tools development.  
1225 **Bioinformatics**, v. 24, n. 21, p. 2534-6, Nov 01 2008. ISSN 1367-4811 (Electronic)  
1226 1367-4803 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18606607> >.  
1227
- 1228 KLEFFMANN, T.; SCHMIDT, J.; SCHAUB, G. A. Attachment of Trypanosoma cruzi epimastigotes  
1229 to hydrophobic substrates and use of this property to separate stages and promote metacyclogenesis.  
1230 **J Eukaryot Microbiol**, v. 45, n. 5, p. 548-55, Sep-Oct 1998. ISSN 1066-5234 (Print)  
1231 1066-5234 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9783457> >.  
1232
- 1233 KOCHER, T. et al. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new  
1234 mass spectrometric method combining the benefits of all. **J Proteome Res**, v. 8, n. 10, p. 4743-52,  
1235 Oct 2009. ISSN 1535-3907 (Electronic)  
1236 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19663507> >.  
1237
- 1238 KRAMER, S.; CARRINGTON, M. Trans-acting proteins regulating mRNA maturation, stability and  
1239 translation in trypanosomatids. **Trends Parasitol**, v. 27, n. 1, p. 23-30, Jan 2011. ISSN 1471-5007  
1240 (Electronic)  
1241 1471-4922 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20609625> >.  
1242
- 1243 KUSTER, B. et al. Scoring proteomes with proteotypic peptide probes. **Nat Rev Mol Cell Biol**, v. 6,  
1244 n. 7, p. 577-83, Jul 2005. ISSN 1471-0072 (Print)  
1245 1471-0072 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15957003> >.  
1246
- 1247 LAM, H. Building and searching tandem mass spectral libraries for peptide identification. **Mol Cell**  
1248 **Proteomics**, v. 10, n. 12, p. R111 008565, Dec 2011. ISSN 1535-9484 (Electronic)  
1249 1535-9476 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21900153> >.  
1250
- 1251 LAM, H.; AEBERSOLD, R. Building and searching tandem mass (MS/MS) spectral libraries for  
1252 peptide identification in proteomics. **Methods**, v. 54, n. 4, p. 424-31, Aug 2011. ISSN 1095-9130  
1253 (Electronic)  
1254 1046-2023 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21277371> >.  
1255
- 1256 LAM, H. et al. Development and validation of a spectral library searching method for peptide  
1257 identification from ms/ms. **Proteomics**, v. 7, n. 5, p. 655-67, Mar 2007. ISSN 1615-9853 (Print)  
1258 1615-9853 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17295354> >.  
1259
- 1260 LAM, H. et al. Building consensus spectral libraries for peptide identification in proteomics. **Nat**  
1261 **Methods**, v. 5, n. 10, p. 873-5, Oct 2008. ISSN 1548-7105 (Electronic)

- 1262 1548-7091 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18806791> >.  
1263
- 1264 LANÇAS, F. M. A cromatografia líquida moderna e a espectrometria de massas: Finalmente  
1265 “compatíveis”? II. A escolha do analisador de massas. **Scientia Chromatographica**, 2013. ISSN  
1266 1984-4433. Disponível em: < <http://dx.doi.org/10.4322/sc.2013.005> >.  
1267
- 1268 LEE, H. J. et al. Comprehensive genome-wide proteomic analysis of human placental tissue for the  
1269 Chromosome-Centric Human Proteome Project. **J Proteome Res**, v. 12, n. 6, p. 2458-66, Jun 07 2013.  
1270 ISSN 1535-3907 (Electronic)  
1271 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23362793> >.  
1272
- 1273 LEE, M. G.; VAN DER PLOEG, L. H. Transcription of protein-coding genes in trypanosomes by  
1274 RNA polymerase I. **Annu Rev Microbiol**, v. 51, p. 463-89, 1997. ISSN 0066-4227 (Print)  
1275 0066-4227 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9343357> >.  
1276
- 1277 LEONARD, G.; SOANES, D. M.; STEVENS, J. R. Resolving the question of trypanosome  
1278 monophyly: a comparative genomics approach using whole genome data sets with low taxon  
1279 sampling. **Infect Genet Evol**, v. 11, n. 5, p. 955-9, Jul 2011. ISSN 1567-7257 (Electronic)  
1280 1567-1348 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21419879> >.  
1281
- 1282 LEVINE, N. D. <strong>The Trypanosomes of Mammals.</strong> A Zoological Monograph. <span  
1283 class="sc">Cecil A. Hoare.</span> Blackwell, Oxford, England, 1972 (U.S. distributor, Davis,  
1284 Philadelphia). xviii, 750 pp. + plates. \$34.50. **Science**, v. 179, n. 4068, p. 60-60, 1973. Disponível  
1285 em: < <http://science.sciencemag.org/content/sci/179/4068/60.full.pdf> >.  
1286
- 1287 LEWIS, M. D. et al. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA  
1288 content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and  
1289 experimental hybrids. **Int J Parasitol**, v. 39, n. 12, p. 1305-17, Oct 2009. ISSN 1879-0135 (Electronic)  
1290 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19393242> >.  
1291
- 1292 LIMA, L. et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and  
1293 phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). **Acta  
1294 Trop**, v. 151, p. 166-77, Nov 2015. ISSN 1873-6254 (Electronic)  
1295 0001-706X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26200788> >.  
1296
- 1297 LIMA, L. et al. New insights into the evolution of the *Trypanosoma cruzi* clade provided by a new  
1298 trypanosome species tightly linked to Neotropical Pteronotus bats and related to an Australian lineage  
1299 of trypanosomes. **Parasit Vectors**, v. 8, p. 657, Dec 23 2015. ISSN 1756-3305 (Electronic)  
1300 1756-3305 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26701154> >.  
1301
- 1302 LIMA, L. et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes  
1303 in *Trypanosoma cruzi*, *T. cruzi*-like and other trypanosome species. **PLoS One**, v. 7, n. 6, p. e38385,  
1304 2012. ISSN 1932-6203 (Electronic)  
1305 1932-6203 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22685565> >.  
1306
- 1307 LINARDI, D. P. N. A. L. D. M. P. M. **Parasitologia Humana**. Atheneu, 2005.  
1308
- 1309 LINK, A. J. et al. Direct analysis of protein complexes using mass spectrometry. **Nat Biotechnol**, v.  
1310 17, n. 7, p. 676-82, Jul 1999. ISSN 1087-0156 (Print)  
1311 1087-0156 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10404161> >.  
1312



- 1313 LIU, B. et al. Fellowship of the rings: the replication of kinetoplast DNA. **Trends Parasitol**, v. 21,  
1314 n. 8, p. 363-9, Aug 2005. ISSN 1471-4922 (Print)  
1315 1471-4922 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15967722> >.  
1316
- 1317 LLEWELLYN, M. S. et al. Trypanosoma cruzi IIc: phylogenetic and phylogeographic insights from  
1318 sequence and microsatellite analysis and potential impact on emergent Chagas disease. **PLoS Negl**  
1319 **Trop Dis**, v. 3, n. 9, p. e510, Sep 01 2009. ISSN 1935-2735 (Electronic)  
1320 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19721699> >.  
1321
- 1322 LUKES, J. et al. Evolution of parasitism in kinetoplastid flagellates. **Mol Biochem Parasitol**, v. 195,  
1323 n. 2, p. 115-22, Jul 2014. ISSN 1872-9428 (Electronic)  
1324 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24893339> >.  
1325
- 1326 MACEDO, A. M. et al. Trypanosoma cruzi: genetic structure of populations and relevance of genetic  
1327 variability to the pathogenesis of chagas disease. **Mem Inst Oswaldo Cruz**, v. 99, n. 1, p. 1-12, Feb  
1328 2004. ISSN 0074-0276 (Print)  
1329 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15057339> >.  
1330
- 1331 MACEDO, A. M. et al. DNA fingerprints: a tool for identification and determination of the  
1332 relationships between species and strains of Leishmania. **Mol Biochem Parasitol**, v. 53, n. 1-2, p.  
1333 63-70, Jul 1992a. ISSN 0166-6851 (Print)  
1334 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1501645> >.  
1335
- 1336 \_\_\_\_\_. DNA fingerprints: a tool for identification and determination of the relationships between  
1337 species and strains of Leishmania. **Molecular and Biochemical Parasitology**, v. 53, n. 1-2, p. 63-70,  
1338 Jul 1992b. ISSN 0166-6851 (Print)  
1339 0166-6851 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/1501645> >.  
1340
- 1341 MAGALHÃES, J. B.; ANDRADE, S. G. Estudo do comportamento de cepas de Trypanosoma cruzi  
1342 após passagem em diferentes espécies de triatomíneos. **Revista da Sociedade Brasileira de**  
1343 **Medicina Tropical**, v. 24, p. 209-216, 1991. ISSN 0037-8682. Disponível em: <  
1344 [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0037-86821991000400002&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86821991000400002&nrm=iso) >.  
1345
- 1346 MAKAROV, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of  
1347 mass analysis. **Anal Chem**, v. 72, n. 6, p. 1156-62, Mar 15 2000. ISSN 1520-6882 (Electronic)  
1348 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10740853> >.  
1349
- 1350 MAKAROV, A.; SCIGELOVA, M. Coupling liquid chromatography to Orbitrap mass spectrometry.  
1351 **J Chromatogr A**, v. 1217, n. 25, p. 3938-45, Jun 18 2010. ISSN 1873-3778 (Electronic)  
1352 0021-9673 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20299023> >.  
1353
- 1354 MALLICK, P.; KUSTER, B. Proteomics: a pragmatic perspective. **Nat Biotechnol**, v. 28, n. 7, p.  
1355 695-709, Jul 2010. ISSN 1546-1696 (Electronic)  
1356 1087-0156 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20622844> >.  
1357
- 1358 MANN, M.; MENG, C. K.; FENN, J. B. Interpreting mass spectra of multiply charged ions.  
1359 **Analytical Chemistry**, v. 61, n. 15, p. 1702-1708, 1989/08/01 1989. ISSN 0003-2700. Disponível  
1360 em: < <http://pubs.acs.org/doi/abs/10.1021/ac00190a023> >.  
1361
- 1362 MARCILI, A. et al. A new genotype of Trypanosoma cruzi associated with bats evidenced by  
1363 phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based  
1364 on ITS1 rDNA. **Parasitology**, v. 136, n. 6, p. 641-55, May 2009. ISSN 1469-8161 (Electronic)

- 1365 0031-1820 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19368741> >.  
1366
- 1367 MARTENS, L. et al. PRIDE: the proteomics identifications database. **Proteomics**, v. 5, n. 13, p.  
1368 3537-45, Aug 2005. ISSN 1615-9853 (Print)  
1369 1615-9853 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16041671> >.  
1370
- 1371 MARTINEZ-CALVILLO, S. et al. Transcription of *Leishmania major* Friedlin chromosome 1  
1372 initiates in both directions within a single region. **Mol Cell**, v. 11, n. 5, p. 1291-9, May 2003. ISSN  
1373 1097-2765 (Print)  
1374 1097-2765 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12769852> >.  
1375
- 1376 MARTINS-MELO, F. R. et al. Mortality due to Chagas disease in Brazil from 1979 to 2009: trends  
1377 and regional differences. **J Infect Dev Ctries**, v. 6, n. 11, p. 817-24, Nov 26 2012. ISSN 1972-2680  
1378 (Electronic)  
1379 1972-2680 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23277508> >.  
1380
- 1381 \_\_\_\_\_. Prevalence of Chagas disease in Brazil: a systematic review and meta-analysis. **Acta Trop**,  
1382 v. 130, p. 167-74, Feb 2014. ISSN 1873-6254 (Electronic)  
1383 0001-706X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24139912> >.  
1384
- 1385 MARTINS-MELO, F. R. et al. Multiple causes of death related to Chagas' disease in Brazil, 1999 to  
1386 2007. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 45, p. 591-596, 2012. ISSN 0037-  
1387 8682. Disponível em: < [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0037-  
1388 86822012000500010&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86822012000500010&nrm=iso) >.  
1389
- 1390 MASLOV, D. A. et al. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be  
1391 revealed. **Trends Parasitol**, v. 29, n. 1, p. 43-52, Jan 2013. ISSN 1471-5007 (Electronic)  
1392 1471-4922 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23246083> >.  
1393
- 1394 MAZZARINO, M. et al. Mass spectrometric characterization of tamoxifene metabolites in human  
1395 urine utilizing different scan parameters on liquid chromatography/tandem mass spectrometry. **Rapid**  
1396 **Commun Mass Spectrom**, v. 24, n. 6, p. 749-60, Mar 2010. ISSN 1097-0231 (Electronic)  
1397 0951-4198 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20187079> >.  
1398
- 1399 MCANDREW, M. et al. Testing Promoter Activity in the Trypanosome Genome: Isolation of a  
1400 Metacyclic-Type VSG Promoter, and Unexpected Insights into RNA Polymerase II Transcription.  
1401 **Experimental Parasitology**, v. 90, n. 1, p. 65-76, 1998/09/01/ 1998. ISSN 0014-4894. Disponível  
1402 em: < <http://www.sciencedirect.com/science/article/pii/S001448949894317X> >.  
1403
- 1404 MELO, R. C.; BRENER, Z. Tissue tropism of different *Trypanosoma cruzi* strains. **J Parasitol**, v. 64,  
1405 n. 3, p. 475-82, Jun 1978. ISSN 0022-3395 (Print)  
1406 0022-3395 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/96243> >.  
1407
- 1408 MESSENGER, L. A.; MILES, M. A.; BERN, C. Between a bug and a hard place: *Trypanosoma cruzi*  
1409 genetic diversity and the clinical outcomes of Chagas disease. **Expert Rev Anti Infect Ther**, v. 13,  
1410 n. 8, p. 995-1029, Aug 2015. ISSN 1744-8336 (Electronic)  
1411 1478-7210 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26162928> >.  
1412
- 1413 MILES, M. A. et al. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause  
1414 Venezuelan and Brazilian forms of Chagas' disease? **Lancet**, v. 1, n. 8234, p. 1338-40, Jun 20 1981.  
1415 ISSN 0140-6736 (Print)  
1416 0140-6736 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6113312> >.

- 1417  
1418 MILES, M. A. et al. Further enzymic characters of *Trypanosoma cruzi* and their evaluation for strain  
1419 identification. **Trans R Soc Trop Med Hyg**, v. 74, n. 2, p. 221-37, 1980. ISSN 0035-9203 (Print)  
1420 0035-9203 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6992358> >.  
1421
- 1422 MILES, M. A. et al. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients  
1423 with Chagas' disease in Amazonian Brazil. **Nature**, v. 272, n. 5656, p. 819-21, Apr 27 1978. ISSN  
1424 0028-0836 (Print)  
1425 0028-0836 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/417267> >.  
1426
- 1427 MILES, M. A. et al. The identification by isoenzyme patterns of two distinct strain-groups of  
1428 *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. **Trans R Soc Trop Med Hyg**,  
1429 v. 71, n. 3, p. 217-25, 1977. ISSN 0035-9203 (Print)  
1430 0035-9203 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/407674> >.  
1431
- 1432 MISCHERIKOW, N. et al. Gaining efficiency by parallel quantification and identification of iTRAQ-  
1433 labeled peptides using HCD and decision tree guided CID/ETD on an LTQ Orbitrap. **Analyst**, v. 135,  
1434 n. 10, p. 2643-52, Oct 2010. ISSN 1364-5528 (Electronic)  
1435 0003-2654 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20714520> >.  
1436
- 1437 MOREIRA, D.; LOPEZ-GARCIA, P.; VICKERMAN, K. An updated view of kinetoplastid  
1438 phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of  
1439 the class Kinetoplastea. **Int J Syst Evol Microbiol**, v. 54, n. Pt 5, p. 1861-75, Sep 2004. ISSN 1466-  
1440 5026 (Print)  
1441 1466-5026 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15388756> >.  
1442
- 1443 MOREIRA, L. M. **Ciências Genômicas: Fundamentos e Aplicações**. Editora Cubo, 2015. ISBN  
1444 978-85-89265-22-5. Disponível em: < <http://moreiralab.net> >.  
1445
- 1446 MOREL, C. et al. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of  
1447 restriction endonuclease products of kinetoplast DNA minicircles. **Proc Natl Acad Sci U S A**, v. 77,  
1448 n. 11, p. 6810-4, Nov 1980. ISSN 0027-8424 (Print)  
1449 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6256762> >.  
1450
- 1451 NDAO, M. Diagnosis of parasitic diseases: old and new approaches. **Hindawi Publishing**  
1452 **Corporation**, p. 15, 2009.  
1453
- 1454 NESVIZHSKII, A. I. A survey of computational methods and error rate estimation procedures for  
1455 peptide and protein identification in shotgun proteomics. **J Proteomics**, v. 73, n. 11, p. 2092-123, Oct  
1456 10 2010. ISSN 1876-7737 (Electronic)  
1457 1874-3919 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20816881> >.  
1458
- 1459 NOGUEIRA, F. C.; DOMONT, G. B. Survey of shotgun proteomics. **Methods Mol Biol**, v. 1156, p.  
1460 3-23, 2014. ISSN 1940-6029 (Electronic)  
1461 1064-3745 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24791978> >.  
1462
- 1463 OCHAYA, S. **Comparative genomics and molecular characterization of N-alpha**  
1464 **Acetyltransferase in Trypanosomes for drug target identification**. 2013. 57 (Doctoral Theses).  
1465 Dept of Cell and Molecular Biology, Karolinska Institutet  
1466
- 1467 OLSEN, J. V. et al. Higher-energy C-trap dissociation for peptide modification analysis. **Nat**  
1468 **Methods**, v. 4, n. 9, p. 709-12, Sep 2007. ISSN 1548-7091 (Print)

- 1469 1548-7091 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17721543> >.  
1470
- 1471 OLSEN, O. W. **Animal Parasites: Their Life Cycles and Ecology**. Dover Publications 1986. 28  
1472 ISBN ISBN-10: 0486651266; ISBN-13: 978-048665126.  
1473
- 1474 ONDER, O. et al. Identifying sources of tick blood meals using unidentified tandem mass spectral  
1475 libraries. **Nat Commun**, v. 4, p. 1746, 2013. ISSN 2041-1723 (Electronic)  
1476 2041-1723 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23612287> >.  
1477
- 1478 OSORIO, L. et al. Virulence factors of *Trypanosoma cruzi*: who is who? **Microbes Infect**, v. 14, n.  
1479 15, p. 1390-402, Dec 2012. ISSN 1769-714X (Electronic)  
1480 1286-4579 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23006853> >.  
1481
- 1482 OSORIO, L. et al. Virulence factors of *Trypanosoma cruzi*: who is who? **Microbes and Infection**,  
1483 v. 14, n. 15, p. 1390-1402, 2012/12/01/ 2012. ISSN 1286-4579. Disponível em: <  
1484 <http://www.sciencedirect.com/science/article/pii/S1286457912002201> >.  
1485
- 1486 PABA, J. et al. Proteomic analysis of *Trypanosoma cruzi* developmental stages using isotope-coded  
1487 affinity tag reagents. **J Proteome Res**, v. 3, n. 3, p. 517-24, May-Jun 2004. ISSN 1535-3893 (Print)  
1488 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15253433> >.  
1489
- 1490 PABA, J. et al. Proteomic analysis of the human pathogen *Trypanosoma cruzi*. **Proteomics**, v. 4, n.  
1491 4, p. 1052-9, Apr 2004. ISSN 1615-9853 (Print)  
1492 1615-9853 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15048986> >.  
1493
- 1494 PAHO. **Chagas disease**: Pan American Health Organization 2017.  
1495
- 1496 PALMISANO, G. et al. A novel method for the simultaneous enrichment, identification, and  
1497 quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse  
1498 brain development. **Mol Cell Proteomics**, v. 11, n. 11, p. 1191-202, Nov 2012a. ISSN 1535-9484  
1499 (Electronic)  
1500 1535-9476 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22843994> >.  
1501
- 1502 \_\_\_\_\_. A novel method for the simultaneous enrichment, identification, and quantification of  
1503 phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain  
1504 development. **Molecular Cell Proteomics**, v. 11, n. 11, p. 1191-202, Nov 2012b. ISSN 1535-9484  
1505 (Electronic)  
1506 1535-9476 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22843994> >.  
1507
- 1508 PARODI-TALICE, A. et al. Proteomic analysis of metacyclic trypomastigotes undergoing  
1509 *Trypanosoma cruzi* metacyclogenesis. **J Mass Spectrom**, v. 42, n. 11, p. 1422-32, Nov 2007. ISSN  
1510 1076-5174 (Print)  
1511 1076-5174 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17960573> >.  
1512
- 1513 PARSONS, M. et al. Trypanosome mRNAs share a common 5' spliced leader sequence. **Cell**, v. 38,  
1514 n. 1, p. 309-16, Aug 1984. ISSN 0092-8674 (Print)  
1515 0092-8674 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6088073> >.  
1516
- 1517 PATTERSON, S. D. Data analysis--the Achilles heel of proteomics. **Nat Biotechnol**, v. 21, n. 3, p.  
1518 221-2, Mar 2003. ISSN 1087-0156 (Print)  
1519 1087-0156 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12610558> >.  
1520

- 1521 PATTERSON, S. D.; AEBERSOLD, R. H. Proteomics: the first decade and beyond. **Nat Genet**, v.  
1522 33 Suppl, p. 311-23, Mar 2003. ISSN 1061-4036 (Print)  
1523 1061-4036 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12610541> >.  
1524
- 1525 PEDRIOLI, P. G. Trans-proteomic pipeline: a pipeline for proteomic analysis. **Methods Mol Biol**, v.  
1526 604, p. 213-38, 2010. ISSN 1940-6029 (Electronic)  
1527 1064-3745 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20013374> >.  
1528
- 1529 PEDRIOLI, P. G. et al. A common open representation of mass spectrometry data and its application  
1530 to proteomics research. **Nat Biotechnol**, v. 22, n. 11, p. 1459-66, Nov 2004. ISSN 1087-0156 (Print)  
1531 1087-0156 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15529173> >.  
1532
- 1533 PEREZ-MORALES, D. et al. Proteomic analysis of *Trypanosoma cruzi* epimastigotes subjected to  
1534 heat shock. **J Biomed Biotechnol**, v. 2012, p. 902803, 2012. ISSN 1110-7251 (Electronic)  
1535 1110-7243 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22287837> >.  
1536
- 1537 PEREZ BRANDAN, C. et al. Knockout of the dhfr-ts gene in *Trypanosoma cruzi* generates  
1538 attenuated parasites able to confer protection against a virulent challenge. **PLoS Negl Trop Dis**, v. 5,  
1539 n. 12, p. e1418, Dec 2011. ISSN 1935-2735 (Electronic)  
1540 1935-2727 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22180798> >.  
1541
- 1542 PERKINS, D. N. et al. Probability-based protein identification by searching sequence databases  
1543 using mass spectrometry data. **Electrophoresis**, v. 20, n. 18, p. 3551-67, Dec 1999. ISSN 0173-0835  
1544 (Print)  
1545 0173-0835 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10612281> >.  
1546
- 1547 PETRY, K.; BALTZ, T.; SCHOTTELIUS, J. Differentiation of *Trypanosoma cruzi*, *T. cruzi*  
1548 *marinkellei*, *T. dionisii* and *T. vespertilionis* by monoclonal antibodies. **Acta Trop**, v. 43, n. 1, p. 5-  
1549 13, Mar 1986. ISSN 0001-706X (Print)  
1550 0001-706X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/2424290> >.  
1551
- 1552 PETRY, K.; VOISIN, P.; BALTZ, T. Complex lipids as common antigens to *Trypanosoma cruzi*, *T.*  
1553 *dionisii*, *T. vespertilionis* and nervous tissue (astrocytes, neurons). **Acta Trop**, v. 44, n. 4, p. 381-6,  
1554 Dec 1987. ISSN 0001-706X (Print)  
1555 0001-706X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/2449808> >.  
1556
- 1557 PIACENZA, L. et al. Enzymes of the antioxidant network as novel determiners of *Trypanosoma*  
1558 *cruzi* virulence. **Int J Parasitol**, v. 39, n. 13, p. 1455-64, Nov 2009. ISSN 1879-0135 (Electronic)  
1559 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19505468> >.  
1560
- 1561 PORCEL, B. M. et al. The streamlined genome of *Phytomonas* spp. relative to human pathogenic  
1562 kinetoplastids reveals a parasite tailored for plants. **PLoS Genet**, v. 10, n. 2, p. e1004007, Feb 2014.  
1563 ISSN 1553-7404 (Electronic)  
1564 1553-7390 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24516393> >.  
1565
- 1566 PRATA, A. Clinical and epidemiological aspects of Chagas disease. **Lancet Infect Dis**, v. 1, n. 2, p.  
1567 92-100, Sep 2001. ISSN 1473-3099 (Print)  
1568 1473-3099 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11871482> >.  
1569
- 1570 PREUSSER, C.; JAE, N.; BINDEREIF, A. mRNA splicing in trypanosomes. **Int J Med Microbiol**,  
1571 v. 302, n. 4-5, p. 221-4, Oct 2012. ISSN 1618-0607 (Electronic)  
1572 1438-4221 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22964417> >.

- 1573  
1574 PUNUKOLLU, G. et al. Clinical aspects of the Chagas' heart disease. **Int J Cardiol**, v. 115, n. 3, p.  
1575 279-83, Feb 14 2007. ISSN 1874-1754 (Electronic)  
1576 0167-5273 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16769134> >.  
1577
- 1578 QUEIROZ, R. M. et al. Cell surface proteome analysis of human-hosted *Trypanosoma cruzi* life  
1579 stages. **J Proteome Res**, v. 13, n. 8, p. 3530-41, Aug 01 2014. ISSN 1535-3907 (Electronic)  
1580 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24978697> >.  
1581
- 1582 QUEIROZ, R. M. et al. Comprehensive proteomic analysis of *Trypanosoma cruzi* epimastigote cell  
1583 surface proteins by two complementary methods. **J Proteome Res**, v. 12, n. 7, p. 3255-63, Jul 05  
1584 2013. ISSN 1535-3907 (Electronic)  
1585 1535-3893 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23682730> >.  
1586
- 1587 RABILLOUD, T. Membrane proteins and proteomics: love is possible, but so difficult.  
1588 **Electrophoresis**, v. 30 Suppl 1, p. S174-80, Jun 2009. ISSN 1522-2683 (Electronic)  
1589 0173-0835 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19517508> >.  
1590
- 1591 RABILLOUD, T. et al. Two-dimensional gel electrophoresis in proteomics: Past, present and future.  
1592 **J Proteomics**, v. 73, n. 11, p. 2064-77, Oct 10 2010. ISSN 1876-7737 (Electronic)  
1593 1874-3919 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20685252> >.  
1594
- 1595 RASSI A JR, A. R., JOSÉ ANTONIO MARIN-NETO. Chagas disease. **Lancet**, v. 375, p. 1388-402,  
1596 2010.  
1597
- 1598 RASSI, A., JR.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **Lancet**, v. 375, n. 9723, p. 1388-  
1599 402, Apr 17 2010. ISSN 1474-547X (Electronic)  
1600 0140-6736 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20399979> >.  
1601
- 1602 RODRIGUEZ, A. et al. Lysosomes behave as Ca<sup>2+</sup>-regulated exocytic vesicles in fibroblasts and  
1603 epithelial cells. **J Cell Biol**, v. 137, n. 1, p. 93-104, Apr 07 1997. ISSN 0021-9525 (Print)  
1604 0021-9525 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9105039> >.  
1605
- 1606 RODRIQUES COURA, J.; DE CASTRO, S. L. A critical review on Chagas disease chemotherapy.  
1607 **Mem Inst Oswaldo Cruz**, v. 97, n. 1, p. 3-24, Jan 2002. ISSN 0074-0276 (Print)  
1608 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11992141> >.  
1609
- 1610 ROEPSTORFF, P.; FOHLMAN, J. Proposal for a common nomenclature for sequence ions in mass  
1611 spectra of peptides. **Biomed Mass Spectrom**, v. 11, n. 11, p. 601, Nov 1984. ISSN 0306-042X (Print)  
1612 0306-042X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6525415> >.  
1613
- 1614 ROUT, M. P.; FIELD, M. C. Isolation and characterization of subnuclear compartments from  
1615 *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. **J Biol Chem**, v.  
1616 276, n. 41, p. 38261-71, Oct 12 2001. ISSN 0021-9258 (Print)  
1617 0021-9258 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11477078> >.  
1618
- 1619 RUDENKO, G. et al. A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene  
1620 expression site can be efficiently switched on and off in *T. brucei*. **Cell**, v. 83, n. 4, p. 547-553,  
1621 1995/11/17/ 1995. ISSN 0092-8674. Disponível em: <  
1622 <http://www.sciencedirect.com/science/article/pii/0092867495900942> >.  
1623

- 1624 RUDENKO, G. et al. RNA polymerase I can mediate expression of CAT and neo protein-coding  
1625 genes in *Trypanosoma brucei*. **EMBO J**, v. 10, n. 11, p. 3387-97, Nov 1991. ISSN 0261-4189 (Print)  
1626 0261-4189 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1915299> >.  
1627
- 1628 SADYGOV, R. G.; COCIORVA, D.; YATES, J. R., 3RD. Large-scale database searching using  
1629 tandem mass spectra: looking up the answer in the back of the book. **Nat Methods**, v. 1, n. 3, p. 195-  
1630 202, Dec 2004. ISSN 1548-7091 (Print)  
1631 1548-7091 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15789030> >.  
1632
- 1633 SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed  
1634 synthesis with DNA polymerase. **J Mol Biol**, v. 94, n. 3, p. 441-8, May 25 1975. ISSN 0022-2836  
1635 (Print)  
1636 0022-2836 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1100841> >.  
1637
- 1638 SAÚDE, M. D. **Boletim Epidemiológico**. Ministério da Saúde - Secretaria de Vigilância em Saúde  
1639 Ministério da Saúde - Secretaria de Vigilância em Saúde Ministério da Saúde. 46 2015.  
1640
- 1641 SCOTT, D. A.; DOCAMPO, R. Characterization of isolated acidocalcisomes of *Trypanosoma cruzi*.  
1642 **J Biol Chem**, v. 275, n. 31, p. 24215-21, Aug 04 2000. ISSN 0021-9258 (Print)  
1643 0021-9258 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10816577> >.  
1644
- 1645 SHAO, W. et al. A peptide identification-free, genome sequence-independent shotgun proteomics  
1646 workflow for strain-level bacterial differentiation. **Sci Rep**, v. 5, p. 14337, Sep 23 2015. ISSN 2045-  
1647 2322 (Electronic)  
1648 2045-2322 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26395646> >.  
1649
- 1650 SILVA, A. R. F. et al. DiagnoProt: a tool for discovery of new molecules by mass spectrometry.  
1651 **Bioinformatics**, v. 33, n. 12, p. 1883-1885, Jun 15 2017a. ISSN 1367-4811 (Electronic)  
1652 1367-4803 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28186229> >.  
1653
- 1654 \_\_\_\_\_. DiagnoProt: a tool for discovery of new molecules by mass spectrometry. **Bioinformatics**,  
1655 Feb 10 2017b. ISSN 1367-4811 (Electronic)  
1656 1367-4803 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28186229> >.  
1657
- 1658 SIMPSON, A. G. B.; ROGER, A. J. Protein phylogenies robustly resolve the deep-level relationships  
1659 within Euglenozoa. **Molecular Phylogenetics and Evolution**, v. 30, n. 1, p. 201-212, 2004/01/01/  
1660 2004. ISSN 1055-7903. Disponível em: <  
1661 <http://www.sciencedirect.com/science/article/pii/S1055790303001775> >.  
1662
- 1663 SINGHAL, N. et al. MALDI-TOF mass spectrometry: an emerging technology for microbial  
1664 identification and diagnosis. **Frontiers in Microbiology**, v. 6, p. 791, 08/05  
1665 03/17/received  
1666 07/21/accepted 2015. ISSN 1664-302X. Disponível em: <  
1667 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525378/> >.  
1668
- 1669 STANISLAUS, R. et al. An open-source representation for 2-DE-centric proteomics and support  
1670 infrastructure for data storage and analysis. **BMC Bioinformatics**, v. 9, p. 4, Jan 07 2008. ISSN 1471-  
1671 2105 (Electronic)  
1672 1471-2105 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18179696> >.  
1673
- 1674 STEEN, H.; MANN, M. The ABC's (and XYZ's) of peptide sequencing. **Nat Rev Mol Cell Biol**, v.  
1675 5, n. 9, p. 699-711, Sep 2004. ISSN 1471-0072 (Print)

- 1676 1471-0072 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15340378> >.  
1677
- 1678 STEIN, S. E.; SCOTT, D. R. Optimization and testing of mass spectral library search algorithms for  
1679 compound identification. **Journal of the American Society for Mass Spectrometry**, v. 5, n. 9, p.  
1680 859-866, 1994/09/01 1994. ISSN 1879-1123. Disponível em: < [https://doi.org/10.1016/1044-](https://doi.org/10.1016/1044-0305(94)87009-8)  
1681 [0305\(94\)87009-8](https://doi.org/10.1016/1044-0305(94)87009-8) >.  
1682
- 1683 STEIN, S. E.; SCOTT, D. R. Optimization and testing of mass spectral library search algorithms for  
1684 compound identification. **J Am Soc Mass Spectrom**, v. 5, n. 9, p. 859-66, Sep 1994. ISSN 1044-  
1685 0305 (Print)  
1686 1044-0305 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24222034> >.  
1687
- 1688 STEINDEL, M. et al. Random amplified polymorphic DNA analysis of Trypanosoma cruzi strains.  
1689 **Mol Biochem Parasitol**, v. 60, n. 1, p. 71-9, Jul 1993. ISSN 0166-6851 (Print)  
1690 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8366896> >.  
1691
- 1692 STEVENS, J. R. et al. The molecular evolution of Trypanosomatidae. **Adv Parasitol**, v. 48, p. 1-56,  
1693 2001. ISSN 0065-308X (Print)  
1694 0065-308X (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11013754> >.  
1695
- 1696 STUART, K. et al. RNA editing in kinetoplastid protozoa. **Microbiol Mol Biol Rev**, v. 61, n. 1, p.  
1697 105-20, Mar 1997. ISSN 1092-2172 (Print)  
1698 1092-2172 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9106367> >.  
1699
- 1700 STURM, N. R. et al. Evidence for multiple hybrid groups in Trypanosoma cruzi. **Int J Parasitol**, v.  
1701 33, n. 3, p. 269-79, Mar 2003. ISSN 0020-7519 (Print)  
1702 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12670512> >.  
1703
- 1704 SYKA, J. E. et al. Peptide and protein sequence analysis by electron transfer dissociation mass  
1705 spectrometry. **Proc Natl Acad Sci U S A**, v. 101, n. 26, p. 9528-33, Jun 29 2004. ISSN 0027-8424  
1706 (Print)  
1707 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15210983> >.  
1708
- 1709 SZKLARCZYK, D. et al. STRING v10: protein-protein interaction networks, integrated over the  
1710 tree of life. **Nucleic Acids Res**, v. 43, n. Database issue, p. D447-52, Jan 2015. ISSN 1362-4962  
1711 (Electronic)  
1712 0305-1048 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25352553> >.  
1713
- 1714 TARROUX, P.; VINCENS, P.; RABILLOUD, T. HERMeS: A second generation approach to the  
1715 automatic analysis of two-dimensional electrophoresis gels. Part V: Data analysis.  
1716 **ELECTROPHORESIS**, 1987. Disponível em: < <http://dx.doi.org/10.1002/elps.1150080404> >.  
1717
- 1718 TELLERIA, J. et al. Predominant clonal evolution leads to a close parity between gene expression  
1719 profiles and subspecific phylogeny in Trypanosoma cruzi. **Mol Biochem Parasitol**, v. 137, n. 1, p.  
1720 133-41, Sep 2004. ISSN 0166-6851 (Print)  
1721 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15279959> >.  
1722
- 1723 TELLERIA, J. et al. Phylogenetic character mapping of proteomic diversity shows high correlation  
1724 with subspecific phylogenetic diversity in Trypanosoma cruzi. **Proc Natl Acad Sci U S A**, v. 107, n.  
1725 47, p. 20411-6, Nov 23 2010. ISSN 1091-6490 (Electronic)  
1726 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21059959> >.  
1727



- 1728 TIBAYRENC, M.; AYALA, F. J. Isozyme Variability in *Trypanosoma Cruzi*, the Agent of Chagas'  
1729 Disease: Genetical, Taxonomical, and Epidemiological Significance. **Evolution**, v. 42, n. 2, p. 277-  
1730 292, Mar 1988. ISSN 1558-5646 (Electronic)  
1731 0014-3820 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28567853> >.  
1732
- 1733 TIBAYRENC, M. et al. Genetic arguments against actual mating in *Trypanosoma cruzi*. Taxonomic  
1734 implications. **Comptes Rendus des Seances de l'Academie des Sciences - Series III**, v. 293, n. 3,  
1735 p. 207-209, 1981. Disponível em: < [https://www.scopus.com/inward/record.uri?eid=2-s2.0-  
1736 0019869579&partnerID=40&md5=ee8cd9c9f908d8f0c14a4f4693895efc](https://www.scopus.com/inward/record.uri?eid=2-s2.0-0019869579&partnerID=40&md5=ee8cd9c9f908d8f0c14a4f4693895efc) >.  
1737
- 1738 TIBAYRENC, M.; LE RAY, D. General classification of the isoenzymic strains of *Trypanosoma*  
1739 (*Schizotrypanum*) *cruzi* and comparison with *T. (S.) C. marinkellei* and *T. (Herpetosoma) rangeli*.  
1740 **Ann Soc Belg Med Trop**, v. 64, n. 3, p. 239-48, Sep 1984. ISSN 0772-4128 (Print)  
1741 0772-4128 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6391397> >.  
1742
- 1743 TIBAYRENC, M. et al. Genetic characterization of six parasitic protozoa: parity between random-  
1744 primer DNA typing and multilocus enzyme electrophoresis. **Proc Natl Acad Sci U S A**, v. 90, n. 4, p.  
1745 1335-9, Feb 15 1993a. ISSN 0027-8424 (Print)  
1746 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8433991> >.  
1747
- 1748 \_\_\_\_\_. Genetic characterization of six parasitic protozoa: parity between random-primer DNA  
1749 typing and multilocus enzyme electrophoresis. **Proceedings of the National Academy of Sciences**,  
1750 v. 90, n. 4, p. 1335-9, Feb 15 1993b. ISSN 0027-8424 (Print)  
1751 0027-8424 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8433991> >.  
1752
- 1753 TIBAYRENC, M. et al. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have  
1754 a complex multiclonal structure. **Proc Natl Acad Sci U S A**, v. 83, n. 1, p. 115-9, Jan 1986. ISSN  
1755 0027-8424 (Print)  
1756 0027-8424 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/3510428> >.  
1757
- 1758 TYANOVA, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics  
1759 data. **Nat Methods**, v. 13, n. 9, p. 731-40, Sep 2016. ISSN 1548-7105 (Electronic)  
1760 1548-7091 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27348712> >.  
1761
- 1762 TYLER, K. M.; ENGMAN, D. M. Flagellar elongation induced by glucose limitation is preadaptive  
1763 for *Trypanosoma cruzi* differentiation. **Cell Motil Cytoskeleton**, v. 46, n. 4, p. 269-78, Aug 2000.  
1764 ISSN 0886-1544 (Print)  
1765 0886-1544 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10962481> >.  
1766
- 1767 \_\_\_\_\_. The life cycle of *Trypanosoma cruzi* revisited. **Int J Parasitol**, v. 31, n. 5-6, p. 472-81, May  
1768 01 2001. ISSN 0020-7519 (Print)  
1769 0020-7519 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11334932> >.  
1770
- 1771 VICKERMAN, K. **The diversity of the kinetoplastid flagellates. In Biology of the Kinetoplastida.**  
1772 Academic Press, 1976.  
1773
- 1774 VICTORA, C. G. et al. Health conditions and health-policy innovations in Brazil: the way forward.  
1775 **Lancet**, v. 377, n. 9782, p. 2042-53, Jun 11 2011. ISSN 1474-547X (Electronic)  
1776 0140-6736 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21561659> >.  
1777
- 1778 WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat**  
1779 **Rev Genet**, v. 10, n. 1, p. 57-63, Jan 2009. ISSN 1471-0064 (Electronic)

- 1780 1471-0056 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19015660> >.  
1781
- 1782 WASINGER, V. C. et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma  
1783 genitalium. **Electrophoresis**, v. 16, n. 7, p. 1090-4, Jul 1995. ISSN 0173-0835 (Print)  
1784 0173-0835 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7498152> >.  
1785
- 1786 WESTENBERGER, S. J. et al. Two hybridization events define the population structure of  
1787 Trypanosoma cruzi. **Genetics**, v. 171, n. 2, p. 527-43, Oct 2005. ISSN 0016-6731 (Print)  
1788 0016-6731 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15998728> >.  
1789
- 1790 WHO. **Chagas disease (american trypanosomiasis)**: World Health Organization 2015.  
1791
- 1792 \_\_\_\_\_. Chagas disease (American trypanosomiasis). 2017. Disponível em: <  
1793 <http://www.who.int/mediacentre/factsheets/fs340/en/> >.  
1794
- 1795 WILKINS, M. R. et al. Two-dimensional gel electrophoresis for proteome projects: the effects of  
1796 protein hydrophobicity and copy number. **Electrophoresis**, v. 19, n. 8-9, p. 1501-5, Jun 1998. ISSN  
1797 0173-0835 (Print)  
1798 0173-0835 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9694302> >.  
1799
- 1800 WOLTERS, D. A.; WASHBURN, M. P.; YATES, J. R., 3RD. An automated multidimensional protein  
1801 identification technology for shotgun proteomics. **Anal Chem**, v. 73, n. 23, p. 5683-90, Dec 01 2001.  
1802 ISSN 0003-2700 (Print)  
1803 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11774908> >.  
1804
- 1805 YAN, S. et al. Characterization of the Leishmania donovani ribosomal RNA promoter. **Mol Biochem**  
1806 **Parasitol**, v. 103, n. 2, p. 197-210, Oct 15 1999. ISSN 0166-6851 (Print)  
1807 0166-6851 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10551363> >.  
1808
- 1809 YATES, J. R., 3RD. Mass spectrometry and the age of the proteome. **J Mass Spectrom**, v. 33, n. 1,  
1810 p. 1-19, Jan 1998. ISSN 1076-5174 (Print)  
1811 1076-5174 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9449829> >.  
1812
- 1813 YATES, J. R., 3RD et al. Method to compare collision-induced dissociation spectra of peptides:  
1814 potential for library searching and subtractive analysis. **Anal Chem**, v. 70, n. 17, p. 3557-65, Sep 01  
1815 1998. ISSN 0003-2700 (Print)  
1816 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9737207> >.  
1817
- 1818 YOSHIDA, N. Surface antigens of metacyclic trypomastigotes of Trypanosoma cruzi. **Infect Immun**,  
1819 v. 40, n. 2, p. 836-9, May 1983. ISSN 0019-9567 (Print)  
1820 0019-9567 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/6341250> >.  
1821
- 1822 ZAGO, M. P. et al. TcI Isolates of Trypanosoma cruzi Exploit the Antioxidant Network for Enhanced  
1823 Intracellular Survival in Macrophages and Virulence in Mice. **Infection and Immunity**, v. 84, n. 6,  
1824 p. 1842-1856, June 1, 2016 2016. Disponível em: < <http://iai.asm.org/content/84/6/1842.abstract> >.  
1825
- 1826 ZHANG, J.; WANG, Y.; LI, S. Deuterium isobaric amine-reactive tags for quantitative proteomics.  
1827 **Anal Chem**, v. 82, n. 18, p. 7588-95, Sep 15 2010. ISSN 1520-6882 (Electronic)  
1828 0003-2700 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20715779> >.  
1829
- 1830 ZHANG, Y. et al. Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. **J Am Soc**  
1831 **Mass Spectrom**, v. 20, n. 8, p. 1425-34, Aug 2009. ISSN 1879-1123 (Electronic)

- 1832 1044-0305 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19403316> >.  
1833
- 1834 ZINGALES, B. **Trypanosoma cruzi: um parasita, dois parasitas ou vários parasitas da doença**  
1835 **de chagas?** 2011a. 44-48.  
1836
- 1837 \_\_\_\_\_. *Trypanosoma cruzi*: um parasita, dois parasitas ou vários parasitas da doença de Chagas?  
1838 **Revista da Biologia**, v. 6b, p. 44-48, 2011b.  
1839
- 1840 ZINGALES, B. et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second  
1841 revision meeting recommends TcI to TcVI. **Mem Inst Oswaldo Cruz**, v. 104, n. 7, p. 1051-4, Nov  
1842 2009a. ISSN 1678-8060 (Electronic)  
1843 0074-0276 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20027478> >.  
1844
- 1845 \_\_\_\_\_. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting  
1846 recommends TcI to TcVI. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051-4, Nov 2009b.  
1847 ISSN 1678-8060 (Electronic)  
1848 0074-0276 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20027478> >.  
1849
- 1850 ZINGALES, B. et al. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale,  
1851 epidemiological relevance and research applications. **Infect Genet Evol**, v. 12, n. 2, p. 240-53, Mar  
1852 2012. ISSN 1567-7257 (Electronic)  
1853 1567-1348 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22226704> >.  
1854
- 1855 ZOMERDIJK, J. C.; KIEFT, R.; BORST, P. Efficient production of functional mRNA mediated by  
1856 RNA polymerase I in *Trypanosoma brucei*. **Nature**, v. 353, n. 6346, p. 772-5, Oct 24 1991. ISSN  
1857 0028-0836 (Print)  
1858 0028-0836 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1658658> >.  
1859
- 1860 ZUBAREV, R. A. Electron-capture dissociation tandem mass spectrometry. **Curr Opin Biotechnol**,  
1861 v. 15, n. 1, p. 12-6, Feb 2004. ISSN 0958-1669 (Print)  
1862 0958-1669 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15102460> >.  
1863  
1864

## APÊNDICE C

**Gilberto Santos de Oliveira**

Curriculum Vitae

---

Sou graduado em biomedicina pela Universidade Estadual de Londrina. Depois, comecei como estagiário e depois como aluno no laboratório GlicoProteômica no Instituto de Ciências Biomédicas da Universidade de São Paulo. Desde que comecei a trabalhar com abordagens proteômicas e, em particular, em espectrometria de massas, percebi as enormes possibilidades desta técnica e especialmente para projetos orientados biologicamente.

O projeto do meu mestrado está focado no desenvolvimento de um método de caracterização para discriminar diferentes cepas de *Trypanosoma cruzi* usando ferramentas de espectrometria de massas e bioinformática. Recentemente, publiquei um trabalho sobre a caracterização do proteoma de *T. cruzi* e suas PTM.

**Formação acadêmica/titulação**

- 2015** Mestrado em Ciências (Biologia da Relação Patógeno-Hospedeiro) pgbmp@icb.usp.br.  
Universidade de São Paulo, USP, Sao Paulo, Brasil  
Orientador: Giuseppe Palmisano  
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico
- 2008 - 2013** Graduação em Biomedicina.  
Universidade Estadual de Londrina, UEL, Londrina, Brasil  
Título: POTENCIAL ANTIFÚNGICO DO COMPOSTO DE MSA EM NANOPARTÍCULAS ALGINATO/QUITOSANA LIBERADORAS DE ÓXIDO NÍTRICO  
Orientador: Luciano Aparecido Panagio

**Formação complementar**

- 2017 - 2017** Extensão universitária em ótica de Programação com Java - Turma 1. (Carga horária: 30h).  
Universidade de São Paulo, USP, Sao Paulo, Brasil
- 2016 - 2016** Extensão universitária em Exploring the NCI-60 Pharmacogenomics Dataset at the CellMiner Platform. (Carga horária: 40h).  
Universidade de São Paulo, USP, Sao Paulo, Brasil
- 2015 - 2015** Curso de curta duração em 1º Mini-Curso de Bioinformática Tips and Tricks: overview de estratégias de. (Carga horária: 8h).  
Universidade de São Paulo, USP, Sao Paulo, Brasil
- 2015 - 2015** Curso de curta duração em The use of Mass Spectrometry for disease screening. (Carga horária: 8h).

Universidade de São Paulo, USP, Sao Paulo, Brasil

- 2012 - 2012** Curso de curta duração em Diversidade Genética em Fungos Endofíticos da Regi. (Carga horária: 14h).  
Universidade Estadual de Maringá, UEM, Maringa, Brasil
- 2012 - 2012** Curso de curta duração em Alimentos Probióticos. (Carga horária: 4h).  
Universidade Estadual de Londrina, UEL, Londrina, Brasil
- 2011 - 2011** Curso de curta duração em 2º Cursos Tópicos de Biologia Computacional. (Carga horária: 22h).

## Atuação profissional

### 1. Universidade Estadual de Londrina - UEL

#### Vínculo institucional

- 2013 - 2013** Vínculo: Monitoria Acadêmica , Enquadramento funcional: Microbiologia Geral E Agrícola , Carga horária: 36, Regime: Parcial

### 2. Universidade de São Paulo - USP

#### Vínculo institucional

- 2015 - Atual** Vínculo: Bolsista , Enquadramento funcional: Mestrando, Regime: Dedicção exclusiva laserjet pro mfp m127fn

## Áreas de atuação

1. Proteômica
2. Bioinformática
3. Espectrometria de Massas
4. Protozoologia Parasitária Humana
5. Micologia

## Produção

### Produção bibliográfica

#### Artigos completos publicados em periódicos

1. **Gilberto Santos de Oliveira**, Rebeca Kawahara, Livia Rosa-Fernandes, Carla Cristi Avila, Marta M. G. Teixeira, Martin R. Larsen and Giuseppe Palmisano. Development of a *Trypanosoma cruzi* Strain Typing Assay using MS2 peptide spectral libraries (Tc-STAMS2).

PLOS Neglected Tropical Diseases. 2017. **SUBMETIDO**

2. **DE OLIVEIRA, GILBERTO SANTOS**; KAWAHARA, REBECA; ROSA-FERNANDES, LIVIA; AVILA, CARLA C.; LARSEN, MARTIN R.; PEREIRA ALVES, JOÃO MARCELO; PALMISANO, GIUSEPPE  
Novel DNA coding regions and protein arginylation reveal unexplored *T. cruzi* proteome and PTMs. INTERNATIONAL JOURNAL OF MASS SPECTROMETRY. , v.x, p.x - , 2016.

3. KIMURA, A. H.; **OLIVEIRA, G. S.**; NISHIO, E. K.; SCANDORIEIRO, S.; SOUZA, P. C.; SCHUROFF, P. A.; MEDEIROS, L. P.; BODNAR, G. C.; SARMIENTO, J. J. P.; GAZAL, L. E. S.; SANTOS, P. M. C.; KOGA, V. L.; CYOIA, P. S.; MOREY, A. T.; TATIBANA, B. T.; NAKAZATO, G.; KOBAYASHI, R. K. T.  
MICROBIOLOGIA PARA O ENSINO MÉDIO E TÉCNICO: CONTRIBUIÇÃO DA EXTENSÃO AO ENSINO E APLICAÇÃO DA CIÊNCIA. Revista Conexão UEPG. , 2013.

## **APRESENTAÇÃO DE TRABALHO E PALESTRA**

1. **GS de Oliveira**; Kawahara, R.; ROSA-FERNANDES, L.; AVILA, C. C.; LARSEN, M. R.; ALVES, J. M. P.; PALMISANO, G.

**NOVEL DNA CODING AND PROTEIN ARGINYLTATION REVEAL UNEXPLORED TRYPANOSOMA CRUZI PROTEOME AND PTMS**, 2016. (Congresso,Apresentação de Trabalho)

2. **OLIVEIRA, G. S.**; NICHIO, B. T. L.; SOUZA, P. C.; BOCATE, K.; ALMEIDA, R. S. C.; DURAN, N.; SEABRA, A.; PANAGIO, L. A.

**ESTUDO DA ATIVIDADE DE NANOPARTÍCULAS DE QUITOSANA LIBERADORAS DE ÓXIDO NÍTRICO CONTRA FUNGOS**, 2013. (Congresso,Apresentação de Trabalho)

3. **OLIVEIRA, G. S.**; NISHIO, E. K.; SARMIENTO, J. J. P.; GAZAL, L. E. S.; OGAKI, M. B.; SCANDORIEIRO, S.; CARDOZO, V. F.; TATIBANA, B. T.; NAKAZATO, G.; KOBAYASHI, R. K. T.

**MICROBIOLOGIA APLICADA AO ENSINO TÉCNICO**, 2013. (Simpósio,Apresentação de Trabalho)

4. **OLIVEIRA, G. S.**

**Potencial Antifúngico de Compostos de Nanopartículas de Alginato/Quitossana**, 2013. (Conferência ou palestra,Apresentação de Trabalho)

5. **OLIVEIRA, G. S.**; NISHIO, E. K.; SARMIENTO, J. J. P.; GAZAL, L. E. S.; OGAKI, M. B.; SCANDORIEIRO, S.; CARDOZO, V. F.; NAKAZATO, G.; TATIBANA, B. T.; KOBAYASHI, R. K. T.

**2º SEMINÁRIO DE EXTENSÃO UNIVERSITÁRIA - 'POR EXTENSO'**, 2013. (Outra,Apresentação de Trabalho)

6. **OLIVEIRA, G. S.**; LONGHI, T.V; CARDOZO, V. M; KOBAYASHI, R. K. T; NAKAZATO, G; GARCIA, S; PANAGIO, L.A

**Avaliação do Efeito de Lactobacillus sp. Na invasão de Salmonella Enterica em Células Hep-**

2, 2011. (Congresso,Apresentação de Trabalho)

7. **OLIVEIRA, G. S.**; Canteri, P; LIONI, L.M; OGATTA, S.F.Y; PANAGIO, L.A  
**Estudo da Atividade de Compostos de Rutênio Liberadores de Óxido Nítrico Contra Fungos,**  
2011. (Congresso,Apresentação de Trabalho)

## **ORGANIZAÇÃO DE EVENTO**

1. PALMISANO, G.; **OLIVEIRA, G. S.**; Kawahara, R.; ANGELI, C. B. Rocha, B. V.;  
**OSID - Omics Sciences in Infectious Diseases, 2017.** (Outro, Organização de evento)

2. **OLIVEIRA, G. S.**  
**3 Congresso Paranaense de Ciências Biomédicas, 2013.** (Congresso, Organização de evento)

3. **OLIVEIRA, G. S.**  
**1º Congresso Paranaense de Ciências Biomédicas, 2011.** (Congresso, Organização de evento)