

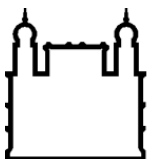
**INSTITUTO CARLOS CHAGAS**  
**MESTRADO EM BIOCÊNCIAS E BIOTECNOLOGIA**

**CARACTERIZAÇÃO DO METILPROTEOMA DE FORMAS EPIMASTIGOTAS DE**  
*Trypanosoma cruzi*

**RAFAEL FOGAÇA DE ALMEIDA**

**CURITIBA**

**2017**



Ministério da Saúde

**FIOCRUZ**  
**Fundação Oswaldo Cruz**

**INSTITUTO CARLOS CHAGAS**  
**Pós-Graduação em Biociências e Biotecnologia**

RAFAEL FOGAÇA DE ALMEIDA

CARACTERIZAÇÃO DO METILPROTEOMA DE FORMAS EPIMASTIGOTAS DE  
*Trypanosoma cruzi*

Dissertação apresentada ao Instituto Carlos Chagas como parte dos requisitos para obtenção do título de Mestre em Biociências e Biotecnologia.

**Orientadora:** Prof. Dra. Lyris Martins Franco de Godoy

**CURITIBA/PR**

2017



Ministério da Saúde

**FIOCRUZ - PARANÁ**  
**Instituto Carlos Chagas**

**Ata da Sessão Pública de exame de dissertação para obtenção do grau de Mestre em  
Bióciências e Biotecnologia.**

Aos 20 dias do mês de fevereiro de dois mil e dezessete, às 14 horas, nas dependências do Instituto Carlos Chagas/Fiocruz Paraná, reuniu-se a Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Biociências e Biotecnologia, composta pelos Professores: Dra. Magda Clara Vieira da Costa-Ribeiro, Dra. Daniela Parada Pavoni e Dra. Paola Fiorani Celedon, com a finalidade de julgar a dissertação do candidato **Rafael Fogaça de Almeida**, intitulada: "**Caracterização do Metilproteoma de *Trypanosoma cruzi***", sob a orientação da Dra. **Lyris Martins Franco de Godoy**, para obtenção do grau de **Mestre em Biociências e Biotecnologia**. O candidato teve até 45 (quarenta e cinco) minutos para a apresentação, e cada examinador teve um tempo máximo de arguição de 30 (trinta) minutos, seguido de 30 (trinta) minutos para resposta do(a) candidato(a) ou de 60 (sessenta) minutos quando houve diálogo na arguição. O desenvolvimento dos trabalhos seguiu o roteiro de sessão de defesa, estabelecido pela Coordenação do Programa, com abertura, condução e encerramento da sessão solene de defesa feito pelo Presidente **Dra. Daniela Parada Pavoni**. Após haver analisado o referido trabalho e arguido a candidata, os membros da banca examinadora deliberaram pela:

- Aprovação da dissertação por unanimidade
- Aprovação somente após satisfazer as exigências de modificações no prazo fixado pela banca (não superior a noventa dias).
- Reprovação da dissertação

Na forma regulamentar foi lavrada a presente ata que é abaixo assinada pelos membros da banca, na ordem acima determinada, e pela candidata.

Curitiba, 20 de Fevereiro de 2017.

Dra. Daniela Parada Pavoni (Presidente) ICC/Fiocruz-Pr 

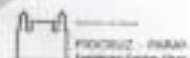
Dra. Magda Clara Vieira da Costa-Ribeiro – UFPR 

Dra. Paola Fiorani Celedon – ICC/Fiocruz-Pr 

Candidato(a): **Rafael Fogaça de Almeida** 

+55(41)3316-3230

Rua Prof. Alcyon Munhoz Mader, 3775 - CIC 81350-010 Curitiba/PR, Brasil



Este trabalho é dedicado aos meus pais.  
Pois não tiveram essa oportunidade.

## **AGRADECIMENTOS**

À Prof. Dra. Lyris, pelos ensinamentos, paciência, ajuda e, acima de tudo, pela oportunidade.

A todos os colegas do Laboratório de Genômica Funcional do Instituto Carlos Chagas.

À Capes e ao Instituto Carlos Chagas, pelo apoio financeiro e institucional.

À banca avaliadora, pelo aceite.

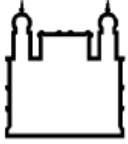
À minha família, por tudo que representa.

A todos que contribuíram nessa trajetória.

À imprevisibilidade da vida, por permitir esse momento.

“Study hard what interests you the most in the most undisciplined,  
irreverent and original manner possible”

**Richard Feynman**



Ministério da Saúde

FIOCRUZ  
Fundação Oswaldo Cruz

## INSTITUTO CARLOS CHAGAS

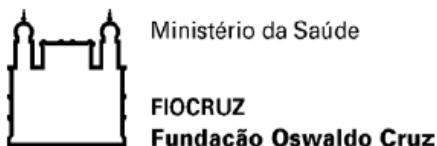
### CARACTERIZAÇÃO DO METILPROTEOMA DE FORMAS EPIMASTIGOTAS DE *Trypanosoma cruzi*

#### RESUMO

#### DISSERTAÇÃO DE MESTRADO

Rafael Fogaça de Almeida

A metilação pós-traducional de proteínas, a qual ocorre em argininas e lisinas, modula diversos processos biológicos, em diferentes níveis de sinalização celular. A proteômica baseada em espectrometria de massas permitiu a identificação de metilargininas em *Trypanosoma brucei*, as quais estão envolvidas em processos fundamentais como metabolismo de RNA, tráfego de proteínas e patogênese. No entanto, a presença e o papel da metilação de proteínas em *Trypanosoma cruzi*, agente etiológico da Doença de Chagas, ainda não foram elucidados. Neste trabalho, foi analisada a presença da maquinaria de metilação/demetilação e caracterizado o metilproteoma de *T. cruzi* em resíduos de arginina e lisina, através de LC-MS/MS. Em epimastigotas, foram identificadas 878 proteínas metiladas e 1336 sítios de metilação (657 metilargininas e 679 metil-lisinas). A metilação em arginina está relacionada a diferentes processos como oxiredução e metabolismo de carboidratos, enquanto a metilação em lisina impacta diretamente a síntese de proteínas. Foi detectada a coocorrência entre metilação e fosforilação em 62 proteínas de *T. cruzi*. Esse trabalho representa a primeira análise, em escala proteômica, do metilproteoma de *T. cruzi*, além de ser o primeiro a caracterizar a metilação em lisinas em tripanossomatídeos. Coletivamente, esses dados informam sobre novos aspectos biológicos fundamentais deste organismo que podem contribuir na identificação de peças chave no processo de adaptação e infecção pelo parasito e, em última instância, indicando possíveis alvos quimioterápicos. Palavras-chave: *Trypanosoma cruzi*. Metilação de proteína. Proteômica. LC-MS/MS.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO CARLOS CHAGAS

### CARACTERIZAÇÃO DO METILPROTEOMA DE FORMAS EPIMASTIGOTAS DE *Trypanosoma cruzi*

#### ABSTRACT

#### DISSERTAÇÃO DE MESTRADO

**Rafael Fogaça de Almeida**

Post-translational methylation of proteins, which occurs in arginines and lysines, modulates several biological processes at different levels of cell signaling. Mass-spectrometry-based proteomics allowed the identification of methylarginines in *Trypanosoma brucei*, which are involved in processes such as RNA metabolism, protein trafficking and pathogenesis. However, the presence and role of protein methylation in *Trypanosoma cruzi*, the etiologic agent of Chagas' disease, has not yet been elucidated. In this work, the presence of the methylation/demethylation machinery in *T. cruzi* was analyzed and the methylproteome of arginine and lysine residues was identified through LC-MS/MS. In epimastigotes, 878 methylated proteins and 1336 methylation sites (657 methylarginines and 679 methyl lysines) were identified. Arginine methylation is related to different processes such as oxireduction and carbohydrate metabolism, while lysine methylation directly impacts the protein synthesis. In addition, the co-occurrence between methylation and phosphorylation was detected in 62 *T. cruzi* proteins. This work represents the first proteomic analysis of *T. cruzi* methylproteome and is the first to characterize lysine methylation in trypanosomatids. Collectively, these data inform about new fundamental biological aspects of this organism that can contribute to the identification of key pieces in the process of adaptation and infection by the parasite and, ultimately, indicating possible candidates for chemotherapeutic targets.

Keywords: *Trypanosoma cruzi*. Protein methylation. Proteomics. LC-MS/MS.



## LISTA DE ILUSTRAÇÕES

Figura 1.1 - Ciclo de vida do <i>T. cruzi</i> .....	8
Figura 1.2 - As PTMs aumentam o nível de complexidade nas células.....	10
Figura 1.3 - Catálise das PMTs a partir de AdoMet e os produtos da metilação de arginina e lisina.....	13
Figura 3.1 - Produção de uma <i>StageTip</i> padrão.....	24
Figura 3.2 - <i>StageTip</i> C18-SCX com 5 porções de cada disco.....	25
Figura 3.3 - Estratégia de concatenação das frações aplicada neste fracionamento.....	31
Figura 3.4 - Modelo do fracionamento do gel gradiente SDS-PAGE.....	32
Figura 4.1 - Desenho experimental empregado na caracterização da maquinaria de metilação e metilproteoma de <i>T. cruzi</i> .....	38
Figura 4.2 – Domínios conservados das PRMTs de <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> ....	41
Figura 4.3 - Alinhamento da região do domínio catalítico das PRMTs de <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> .....	44
Figura 4.4 – Domínios conservados das PKMTs da família DOT1 de <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> .....	47
Figura 4.5 - Alinhamento da região do domínio catalítico das PKMTs da família DOT1 em <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> .....	48
Figura 4.6 - Domínios conservados das demetilases com domínio JmjC em <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> .....	51
Figura 4.7 - Alinhamento da região do domínio catalítico das demetilases de <i>T. cruzi</i> , <i>T. brucei</i> e <i>H. sapiens</i> .....	54
Figura 4.8 - Distribuição Me/Me2/Me3 dos eventos de metilação em epimastigotas de <i>T. cruzi</i> .....	57
Figura 4.9 - Proteínas metiladas de <i>T. cruzi</i> classificadas de acordo com Componente Celular.....	59
Figura 4.10 - Componentes celulares enriquecidos dentre as proteínas metiladas de <i>T. cruzi</i> .....	60

Figura 4.11 – Proteínas metiladas de <i>T. cruzi</i> classificadas de acordo com Processo Biológico.....	61
Figura 4.12 – Processos biológicos enriquecidos dentre as proteínas metiladas de <i>T. cruzi</i> .....	62
Figura 4.13 – Proteínas metiladas de <i>T. cruzi</i> classificadas de acordo com Função Molecular.....	64
Figura 4.14 – Funções moleculares enriquecidas dentre as proteínas metiladas de <i>T. cruzi</i> .....	65
Figura 4.15 – Proteínas metiladas de <i>T. cruzi</i> classificadas de acordo com COG.....	66
Figura 4.16 – Proteínas metiladas de <i>T. cruzi</i> classificadas de acordo com KEGG.....	67
Figura 4.17 – KEGGs enriquecidas dentre as proteínas metiladas de <i>T. cruzi</i> .....	68
Figura 4.18 – Sobreposição do metilproteoma e fosfoproteoma em <i>T. cruzi</i> .....	72
Figura 4.19 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em <i>T. cruzi</i> e categorias funcionais do Cluster de Anotações 1.....	77
Figura 4.20 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em <i>T. cruzi</i> e categorias funcionais do Cluster de Anotações 2.....	78
Figura 4.21 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em <i>T. cruzi</i> e categorias funcionais do Cluster de Anotações 3.....	79

## LISTA DE TABELAS

Tabela 3.1 - Parâmetros utilizados na busca na plataforma MaxQuant.....	35
Tabela 4.1 - TcPRMTs e seus homólogos em <i>T. brucei</i> e <i>H. sapiens</i> .....	39
Tabela 4.2 - TcPKMTs e seus homólogos em <i>T. brucei</i> e <i>H. sapiens</i> .....	45
Tabela 4.3 - Demetilases de <i>T. cruzi</i> e seus homólogos em <i>T. brucei</i> e <i>H. sapiens</i> .....	49
Tabela 4.4 - Proteínas, peptídeos e sítios identificados nos diferentes fracionamentos.....	56
Tabela 4.5 – Categorias funcionais enriquecidas dentre proteínas metiladas em arginina ou lisina em <i>T. cruzi</i> .....	70
Tabela 4.6 - Categorias funcionais enriquecidas no conjunto de proteínas metiladas em arginina em <i>T. brucei</i> .....	71
Tabela 4.7 - Proteínas que apresentam coocorrência de metilação e fosforilação em <i>T. cruzi</i> .....	73
Tabela 4.8 - Clusterização funcional das proteínas que apresentam coocorrência de metilação e fosforilação em <i>T. cruzi</i> .....	76
Tabela S1 - Proteínas metiladas de <i>T. cruzi</i>	
Tabela S2 - Sítios de metilação de <i>T. cruzi</i>	
Tabela S3 - Classificação funcional do metilproteoma de <i>T. cruzi</i>	
Tabela S4 - Enriquecimento funcional metilproteoma do <i>T. cruzi</i>	
Tabela S5 - Clusterização do metilproteoma de <i>T. cruzi</i>	
Tabela S6 - Classificação funcional das proteínas metiladas e fosforiladas em <i>T. cruzi</i>	

## LISTA DE ABREVIATURAS E SIGLAS

ABC - Bicarbonato de Amônio  
ACN - Acetonitrila  
AcOH - Ácido acético  
AdoMet - S-adenosilmetionina  
ATP - Adenosina trifosfato  
BLAST - *Basic Alignment Search Tool*  
BP - *Biological Process*  
CC - *Celullar Component*  
CID - *Collision-Induced Dissociation*  
COG - *Clusters of Orthologous Groups*  
DDA - *Data-Dependent Acquisition*  
DMSO – Dimetilsulfóxido  
DOT - *Disruptor of Telomeric Silencing*  
DTT - Ditioneitol  
E-L - *Esmeraldo-like*  
FA - Formiato de amônio  
HCD - *Higher Collisional Dissociation*  
ESI - *Eletrospray Ionization*  
ETD - *Eletron-transfer dissociation*  
FASP - *Filter-Aided Sample Preparation*  
FDR - *False Discovery Rate*  
GO - *Gene Ontology*  
JHDM - *JmjC-domain-containing histone demethylases*  
Jmj - Jumonji  
kDNA - DNA do cinetoplasto  
KEGG - *Kyoto Encyclopedia of Genes and Genomes*  
Kme - Monometilação de lisina  
Kme2 - Dimetilação de lisina  
Kme3 - Trimetilação de lisina  
LIT – *Liver Infusion Tryptose*

LTQ - *Linear Trap Quadrupole*  
Me - Monometilação  
Me2 - Dimetilação  
Me3 - Trimetilação  
MSA - *Multistage Activation*  
MS - *Mass Spectrometry*  
MS/MS - Espectrometria de massas sequencial  
nanoLC - nanocromatografia líquida  
NCBI - *National Center for Biotechnology Information*  
NH<sub>4</sub>AcO - Acetato de amônio  
NH<sub>4</sub>HCO<sub>3</sub> - Bicarbonato de amônio  
NH<sub>4</sub>OH - Hidróxido de amônio  
NE-L - *non-Esmeraldo-like*  
pb – Pares de base  
PKMT - *Protein Lysine Methyltransferase*  
PMT - Proteína metiltransferase  
PRMT - *Protein Arginine Methyltransferase*  
PTM - *Post-Translational Modification*  
Rme - Monometilação de arginina  
Rme2s - Dimetilação de arginina simétrica  
Rme2a - Dimetilação de arginina assimétrica  
RNAm - RNA mensageiro  
RRM - *RNA Recognition Motif*  
SAM - S-adenosilmetionina  
SCX - *Strong Cation Exchange*  
SET - SU (var), Enhancer de Zeste e Trithorax  
SDS - Sódio Dodecil Sulfato  
SDS-PAGE: *Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis*  
SILAC - *Stable Isotope Labelling by Amino Acids in Cell Culture*  
TFA - *Trifluoroacetic Acid*  
TiO<sub>2</sub> - Dióxido de titânio

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	2
1.1	<b><i>T. cruzi</i> e a Doença de Chagas</b>	5
1.2	<b>Proteínas e suas PTMs</b>	8
1.3	<b>Metilação de proteínas</b>	11
1.3.1	Metilação de arginina (R)	14
1.3.2	Metilação de lisina (K)	14
1.4	<b>Metilproteoma de tripanossomatídeos</b>	15
1.5	<b>Estudo das PTMs por MS</b>	17
1.6	<b>Estudo do metilproteoma por MS</b>	18
<b>2</b>	<b>OBJETIVOS</b>	20
2.1	<b>Objetivo geral</b>	20
2.2	<b>Objetivos específicos</b>	20
<b>3</b>	<b>METODOLOGIA</b>	21
3.1	<b>Busca por maquinaria de metilação/demetilação, domínios conservados e alinhamentos de sequência primária</b>	21
3.2	<b>Cultura de células</b>	22
3.3	<b>Extração e digestão de proteínas em solução</b>	22
3.4	<b>Fracionamentos de peptídeos e proteínas</b>	23
3.4.1	Produção das <i>StageTips</i> ( <i>Stop-and-Go-extraction Tips</i> )	23
3.4.2	Fracionamento por cromatografia de troca iônica - C18-SCX (sal)	24
3.4.3	Fracionamento por cromatografia de troca iônica - C18-SCX (pH)	27
3.4.4	Fracionamento por cromatografia de fase reversa - C18 (pH ácido)	28
3.4.5	Fracionamento por cromatografia de fase reversa - C18 (pH básico)	30
3.4.6	Fracionamento e digestão de proteínas por gel gradiente SDS-PAGE	31
3.5	<b>Análise por nanoLC-MS/MS</b>	33
3.6	<b>Identificação dos peptídeos/proteínas em banco de dados</b>	34
3.7	<b>Classificação funcional do metilproteoma de <i>T. cruzi</i></b>	36
<b>4</b>	<b>RESULTADOS</b>	37
4.1	<b>Análise, <i>in silico</i>, da maquinaria de metilação/demetilação de <i>T. cruzi</i></b>	39
4.1.1	Arginina metiltransferases de <i>T. cruzi</i> (TcPRMTs)	39
4.1.2	Lisina metiltransferases de <i>T. cruzi</i> (TcPKMTs) - Família DOT1	45
4.1.3	Demetilases contendo domínios JmjC de <i>T. cruzi</i>	49
4.2	<b>Caracterização do metilproteoma de formas epimastigotas de <i>T. cruzi</i></b>	55
4.2.1	Análise preliminar, <i>in silico</i> , do metilproteoma de <i>T. cruzi</i>	55

4.2.2	Eficiência de diferentes protocolos de fracionamento na identificação de proteínas metiladas .....	56
4.2.3	O metilproteoma de formas epimastigotas de <i>T. cruzi</i> .....	57
4.2.3.1	Panorama geral da localização e função das proteínas reguladas por metilação em epimastigotas de <i>T. cruzi</i> .....	58
4.2.3.2	Panorama das vias e categorias funcionais envolvidas com os metilproteomas de arginina e lisina em epimastigotas de <i>T. cruzi</i> .....	69
4.2.3.3	Metilação apresenta coocorrência com fosforilação em proteínas de <i>T. cruzi</i> .....	72
<b>5</b>	<b>DISCUSSÃO</b> .....	<b>80</b>
<b>5.1</b>	<b><i>T. cruzi</i> possui uma maquinaria de metilação e demetilação conservada</b> .....	<b>80</b>
<b>5.2</b>	<b>A otimização do preparo de amostras permitiu a identificação em larga escala dos sítios de metilação encontrados em <i>T. cruzi</i></b> .....	<b>82</b>
<b>5.3</b>	<b>A classificação funcional das proteínas metiladas sugere uma potencial influência da metilação em diversos processos biológicos de <i>T. cruzi</i></b> .....	<b>84</b>
<b>5.4</b>	<b>O <i>crosstalk</i> entre metilação e outras modificações em <i>T. cruzi</i></b> .....	<b>86</b>
<b>6</b>	<b>CONCLUSÕES</b> .....	<b>87</b>
	<b>REFERÊNCIAS</b> .....	<b>88</b>
	<b>ANEXOS</b> .....	<b>94</b>
	<b>Anexo 1</b> .....	<b>94</b>

# 1 INTRODUÇÃO

Desde meados da década de 1990, quando foi cunhado o termo “proteoma” para descrever o conjunto total de proteínas expressas pelo genoma da bactéria *Mycoplasma genitalium* (WASINGER et al., 1995), a busca pela identificação em larga escala de proteínas, sua abundância e variação dentro de um proteoma, assim como suas modificações e interações, vem sendo expandida, e isso se deu, principalmente, em associação com o aprimoramento de técnicas envolvendo a proteômica baseada em espectrometria de massas (*Mass Spectrometry*, MS) (AEBERSOLD; MANN, 2003; PATTERSON; AEBERSOLD, 2003; HAN; ASLANIAN; YATES, 2008; GSTAIGER; AEBERSOLD, 2009; AEBERSOLD; MANN, 2016,). A proteômica baseada em MS consiste no amplo estudo de proteínas, permitindo o estudo de milhares de proteínas e suas modificações, a interação entre proteínas e o estudo de organelas, abrangendo muito além do perfil de expressão gênica de uma célula, e devido a esse potencial pode ser considerada uma nova genômica (COX; MANN, 2007).

Dentre as novas áreas possibilitadas pela proteômica baseada em MS, não acessíveis pela genômica, está o estudo em larga escala das modificações pós-traducionais (*Post-Translational Modifications*, PTMs). Dentre as mais conhecidas PTMs estão a fosforilação, metilação, acetilação, ubiquitinação, glicosilação, sumoilação e palmitoilação, entre outras. Estas modificações são catalisadas por enzimas específicas e ocorrem após as proteínas serem traduzidas pelos ribossomos. Assim, como o controle da expressão gênica, as PTMs exercem um papel crucial na regulação das funções celulares, mas de maneira mais precisa e elegante.

Dentre estas modificações, uma das mais comuns em espécies procarióticas e eucarióticas, é a metilação. Apesar de ser conhecida há mais de 50 anos (AMBLER; REES, 1959; MARTINEZ, 1963; MURRAY, 1964; KIM; PAIK, 1965; PAIK; KIM, 1967), a metilação de proteínas ainda é pouco compreendida quanto à sua abrangência e papel final nas células. Seu estudo em larga escala tem sido possível ao passo que os meios analíticos, principalmente os espectrômetros de massas, têm sido aprimorados quanto à sua sensibilidade, assim como com o desenvolvimento de estratégias proteômicas direcionadas ao estudo dessa PTM.



As reações de metilação ocorrem, basicamente, por meio de metiltransferases (PMTs) que utilizam S-adenosilmetionina (AdoMet ou SAM) como molécula doadora de grupos metil (CH<sub>3</sub>), de modo análogo a adenosina trifosfato (ATP) que serve de doador de grupos fosfato (PO<sub>3</sub>) no processo de fosforilação pelas quinases. A metilação ocorre, principalmente, em resíduos de arginina (R), através das arginina metiltransferases (*protein arginine methyltransferases*, PRMTs) e em resíduos de lisina (K), por meio das lisina metiltransferases (*protein lysine methyltransferases*, PKMTs).

Os estudos sobre metilação em proteínas têm demonstrado que, juntamente com a fosforilação, a metilação é uma das principais modificações que controlam funções proteicas, uma vez que tem sido relacionada a processos celulares fundamentais que incluem, por exemplo, a reparação de DNA, regulação transcricional, exportação nuclear de proteínas, *splicing* de RNAm, interação proteína-proteína e proteína-ácido nucleico e transdução de sinal. Estes processos regulados por metilação, em arginina e lisina, têm implicações diretas em diferentes funções biológicas, e juntamente a outros processos de sinalização, contribuem significativamente com a homeostasia celular (BEDFORD; CLARKE, 2009).

A metilação de argininas e lisinas é uma das mais significantes PTMs, responsável por regular as interações celulares em diferentes níveis. Por exemplo, a metilação sítio-específica de lisina é bem estabelecida em proteínas histonas de diferentes organismos, principalmente na regulação da cromatina e sua contribuição ao código de histonas. Além disso, a metilação de proteínas não-histonas e seu papel na regulação biológica também tem sido cada vez mais estudada e demonstrada como tendo grande influência nas interações proteína-proteína e proteína-ácido nucleico (CARLSON; GOZANI, 2014).

De modo geral, o estudo das PTMs tem sido expandido através da proteômica baseada em MS e algumas das estratégias que visam seu estudo se baseiam no enriquecimento das amostras, seja em nível de proteínas e/ou sítios, com a utilização de ligantes específicos a estas PTMs, como anticorpos aprimorados e domínios ligantes, que possuem uma ampla especificidade favorecendo assim sua identificação (CARLSON; GOZANI, 2014).

Outra recente abordagem, de caráter quantitativo, envolve a marcação

metabólica de proteínas com isótopos estáveis de aminoácidos em cultura de células (*Stable Isotope Labelling by Amino Acids in Cell Culture*, SILAC). Por MS, os peptídeos provenientes de diferentes amostras são separados pela diferença de massa introduzida pelo aminoácido marcado. Na metodologia de SILAC, normalmente são utilizados isótopos pesados de lisina e arginina, os quais permitem a quantificação relativa de todas as proteínas entre diferentes amostras. Em uma variação da técnica, denominada *Heavy Methyl* SILAC (ONG; MITTLER; MANN, 2004), células são crescidas em meio contendo metionina marcada com deutério e carbono-13 ( $^{13}\text{CD}_3$ ). Como este aminoácido serve de precursor para produzir grupamentos metil, resíduos metilados marcados também serão incorporados às proteínas, tornando possível a quantificação específica de sítios de metilação. SILAC tem sido uma poderosa estratégia cada vez mais utilizada em proteômica quantitativa nesta última década.

Uma vez que a metilação de proteínas significativamente contribui com a homeostasia celular, a metilação tem sido estudada em alguns parasitos de interesse em saúde pública, como o *Trypanosoma brucei*, que causa a doença do sono em humanos. Em um recente estudo caracterizando o metilproteoma de arginina de *Trypanosoma brucei*, foram identificados 1332 sítios de metilação em arginina distribuídos em 676 proteínas distintas (LOTT et al., 2013). Este trabalho demonstrou que a metilação é abrangente neste organismo, estimando que até 10% do proteoma deste parasito contém metilarginina, presente em diferentes vias nesse tripanossomatídeo. A caracterização, em larga escala, das proteínas metiladas em outros parasitos de interesse epidemiológico, como o *T. cruzi*, agente etiológico da Doença de Chagas, tem, portanto, importante significado biológico uma vez que devido ao seu potencial de elucidar vias importantes sendo impactadas pela metilação de proteínas neste parasito, pode-se encontrar peças chave no processo de infecção deste parasito e ao mesmo tempo candidatos para alvos quimioterápicos.

Assim, este trabalho busca, em escala proteômica, identificar os sítios de metilação em arginina e lisina em *T. cruzi*, cuja ampla caracterização ainda não foi realizada para esse organismo, definindo a amplitude dessa modificação e suas proteínas alvo.

## 1.1 *T. cruzi* e a Doença de Chagas

Os tripanossomatídeos são protozoários flagelados da ordem Kinetoplastida, a qual é caracterizada pela presença do cinetoplasto, uma organela modificada próxima ao corpo basal do flagelo, que contém DNA mitocondrial formando círculos de DNA interligados (VICKERMAN, 1994).

A família Trypanosomatidae apresenta importantes patógenos de interesse médico e veterinário, como *T. cruzi*, *T. brucei* e *Leishmania major* (Trityps). Estes tripanossomatídeos são transmitidos por insetos vetores e apresentam uma variedade de formas morfológicas durante seus ciclos de vida (STEVENS, 2008), os quais envolvem a infecção de vertebrados e invertebrados e, apresentam a necessidade de adaptação fisiológica diante da mudança de ambiente, levando ao complexo processo de diferenciação celular. Essas diferenças morfológicas incluem o ponto de emergência do flagelo do corpo celular, a extensão da membrana ondulante quando existente, a posição do cinetoplasto em relação ao núcleo e ao formato geral da célula. Essas características distinguem as principais formas celulares nos Trityps: epimastigota, tem forma alongada e apresenta cinetoplasto anterior ao núcleo, sua pequena membrana ondulante está disposta na lateral da célula; tripomastigota, também de formato fusiforme, mas com cinetoplasto posterior ao núcleo e o flagelo forma uma longa membrana ondulante livre na porção anterior da célula; e amastigota, de morfologia arredondada e com flagelo curto que não se prolonga da célula (SIQUEIRA, 2004).

O gênero *Trypanosoma* apresenta espécies de ciclo biológico geralmente heteroxêno que são transmitidos por insetos hematófagos e realizam seu ciclo evolutivo em hospedeiros vertebrados onde são encontrados na forma tripomastigota circulante e/ou amastigota. Este gênero encontra-se dividido em duas seções: i) *Stercoraria*, que compreende espécies com desenvolvimento no intestino posterior de seus vetores como o *T. cruzi*, e *T. lewisi*; e ii) *Salivaria*, onde os parasitos tem seu desenvolvimento completo na porção inicial do tubo digestivo e são transmitidos via saliva dos seus vetores, como *T. brucei* e *T. congolense* (GRISARD et al., 1999). Na seção *Stercoraria*, onde se encontra *T. cruzi*, se desenvolve no intestino posterior de seus vetores, pois se adere ao trato digestivo e devido a uma expressão gênica diferencial, passa de uma

forma replicativa e não infectiva (epimastigota) para uma forma não replicativa e infectiva (tripomastigota metacíclico) em um processo de diferenciação chamado metaciclogênese (CONTRERAS; MOREL; GOLDENBERG, 1985).

A Doença de Chagas ou tripanossomíase americana é uma antropozoonose endêmica da América do Sul onde tem relevante importância na saúde pública, no entanto, cada vez mais tem sido encontrada em outras partes do mundo, incluindo países considerados livres da doença, como Japão e Austrália (PEREZ; LYMBERY; THOMPSON, 2015). Aproximadamente 10 mil pessoas morrem todo ano em decorrência das complicações da doença, oito milhões de pessoas estão infectadas e mais de 25 milhões de pessoas estão em risco de infecção em todo o mundo (WHO, 2016).

Descoberta e descrita em 1909 pelo médico brasileiro Carlos Justiniano Ribeiro Chagas (CHAGAS, 1909), a doença que recebe seu sobrenome é ainda uma doença negligenciada, especialmente pelo fato de estar restrita aos países em desenvolvimento onde a disponibilidade de recursos é precária. É causada pelo *T. cruzi*, um protozoário eucarionte, flagelado, digenético, pertencente à classe Mastigophora, ordem Kinetoplastida e família Trypanosomatidae. Possui ciclo silvestre, mas tem se tornado adaptado para um ciclo doméstico, envolvendo seres humanos. Possui mais de 150 hospedeiros mamíferos, incluindo marsupiais, roedores silvestres e domésticos, morcegos, entre outros. No ciclo doméstico, o mais importante reservatório é o próprio homem (DIAS, 1994).

A principal forma de transmissão da Doença de Chagas na América Latina ainda continua sendo pela forma vetorial. Os principais vetores são os insetos *Triatoma infestans*, *Rhodnius prolixus* e *Panstrongylus megistus*, pertencem à ordem Hemiptera, família Reduviidae e subfamília Triatominae (COURA, 2015). Esses triatomíneos, uma vez que são hematófagos, realizam seu repasto sanguíneo em diferentes espécies de vertebrados, dentre elas o homem. Quando o triatomíneo está infectado pelo *T. cruzi*, os parasitos atingem a parte final do intestino posterior do inseto e são liberados pela ampola retal nas fezes do triatomíneo durante o repasto sanguíneo. O homem se infecta quando, instintivamente, ao coçar o local da picada permite que as fezes do barbeiro infectado entrem em contato com a corrente sanguínea, seja através do local

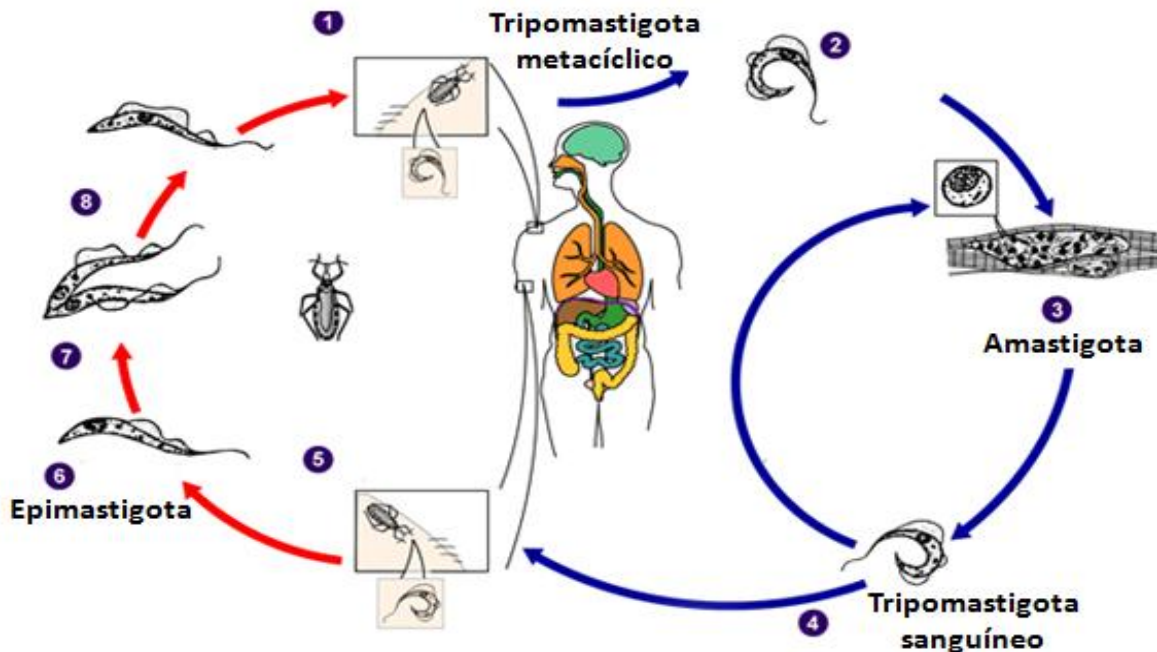
da picada, ou mesmo pelas mucosas ou ferimentos. Ou seja, o ciclo de vida de *T. cruzi* envolve um hospedeiro invertebrado e um hospedeiro vertebrado, sendo transmitida ao homem através do inseto triatomíneo infectado (Figura 1.1). Entretanto, outras vias têm sido representativas recentemente, tais como como transfusão sanguínea e transmissão oral. O *T. cruzi* pode ser transmitido através de alimentos contaminados com as fezes de triatomíneos infectados, por meio de transfusões de sangue utilizando sangue de doadores infectados, passagem de uma mãe infectada para o recém-nascido durante a gravidez, pelo transplante de órgãos que quando os doadores estão infectados, por acidentes de laboratório (WHO, 2016) e ainda por ingestão dos parasitos quando acidentalmente triatomíneos infectados estão presentes durante a preparação de suco de açaí ou cana-de-açúcar (STEINDEL et al., 2008).

Quanto às implicações clínicas, há duas fases. A fase aguda, etapa inicial da doença, é caracterizada por um conjunto de manifestações de intensidade variada. Em geral, neste período há um grande número de parasitos presentes na circulação, devido ao processo de sucessivos ciclos de invasão celular, multiplicação e liberação de parasitos para a corrente sanguínea, essa alta parasitemia sanguínea torna-se aparente entre o 4º e o 40º dia após a infecção, sendo mais evidente do 8º ao 12º dia. Um dos primeiros sinais clínicos agudos é o sinal de Romaña, no olho, ou então o chagoma, na pele, que são decorrentes do processo inflamatório inespecífico no local do sítio de entrada do parasito. Ainda ocorrem sintomas como febre, mal estar, astenia, cefaleia, comprometimento cardíaco; como miocardite, pericardite, endocardite aguda ou taquicardia, podendo ocorrer também lesões discretas no sistema digestório (AMATO NETO; SHIKANAI-YASUDA; AMATO, 1997). No entanto, muitas das pessoas infectadas passam logo após a fase aguda por uma redução na sintomatologia e os níveis de parasitemia caem assim como os anticorpos para os antígenos de *T. cruzi* (CHENG et al., 2007).

A fase crônica pode ser de forma assintomática, ou indeterminada, onde o paciente chagásico pode viver de 20 a 30 anos, levando uma vida normal. No entanto, esta forma pode evoluir para a fase crônica sintomática, na qual a parasitemia torna-se muito baixa, acarretando alterações relacionadas principalmente ao aparelho cardiocirculatório e digestivo. Os principais sintomas desta fase estão relacionados com

doenças do músculo cardíaco, onde os batimentos cardíacos tornam-se arrítmicos, há perda da capacidade de bombeamento cardíaco e cardiomegalia. Outros sintomas característicos desta forma clínica podem ser sintomas no sistema digestivo, como o aumento do esôfago e do intestino grosso, que causam dificuldades de deglutição, engasgos e dor abdominal (NEVES, 2003).

Figura 1.1 - Ciclo de vida do *T. cruzi*.



Fonte: Adaptado de CDC (2015)

O inseto vetor triatomíneo infectado alimenta-se do sangue e libera tripomastigotas em suas fezes, perto do local da ferida. Tripomastigotas entram no hospedeiro vertebrado através da ferida ou através das membranas mucosas (1). Dentro do hospedeiro, os tripomastigotas metacíclicos invadem as células próximas ao local da inoculação, onde se diferenciam em amastigotas intracelulares (2). As amastigotas se multiplicam por divisão binária nas células dos tecidos infectados (3) e se diferenciam em tripomastigotas, em seguida, são liberados na circulação como tripomastigotas na corrente sanguínea (4). Tripomastigotas sanguíneos infectam diferentes células e se transformam em formas amastigotas em novos locais de infecção. O inseto torna-se infectado ao se alimentar de sangue humano ou de outro animal que contenha parasitos circulantes (5). Os tripomastigotas ingeridos se transformam em epimastigotas no intestino médio do vetor (6). Os parasitos se multiplicam e se diferenciam no intestino (7) e se diferenciam em tripomastigotas metacíclicos infectantes no intestino grosso (8).

## 1.2 Proteínas e suas PTMs

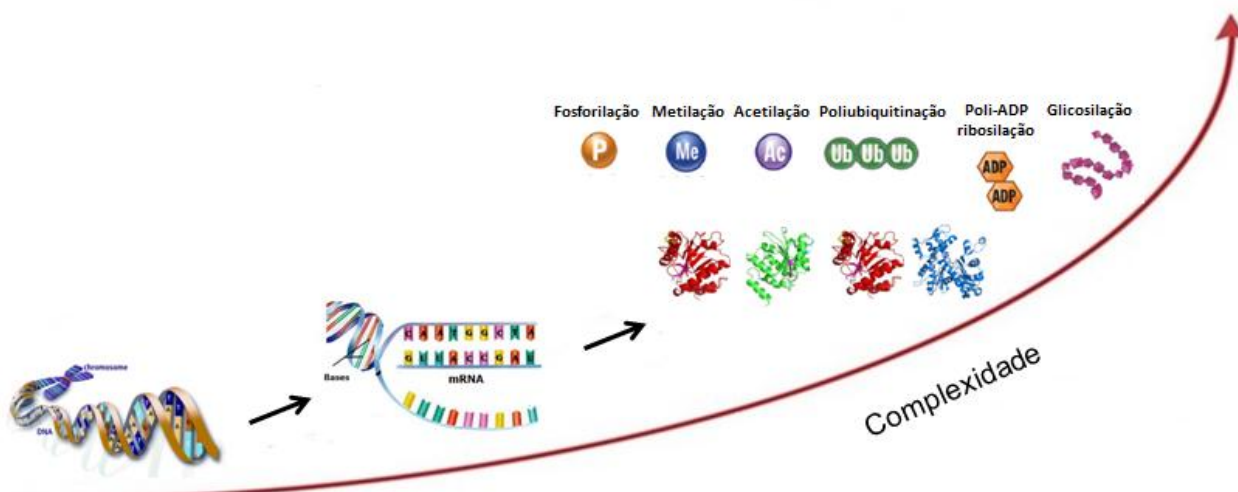
As proteínas são as principais macromoléculas responsáveis pelas funções biológicas em todos os organismos. Sua diversidade funcional está relacionada

essencialmente às diferentes combinações aminoacídicas e às propriedades bioquímicas de cada um dos aminoácidos que compõe sua estrutura. Além disso, estes aminoácidos podem ser modificados pós-traducionalmente, geralmente pela ligação covalente de grupos funcionais às proteínas por enzimas. Estas PTMs determinam o perfil bioquímico do aminoácido e por consequência o comportamento final da proteína. Estas modificações permitem uma variação molecular e funcional além da permitida pelos vinte aminoácidos codificados pela maquinaria celular, aumentando a complexidade do proteoma, pois passa a apresentar formas modificadas e não modificadas das proteínas.

Dentre as mais de 200 PTMs conhecidas, as principais são a fosforilação, metilação, glicosilação, acetilação, ubiquitinação, que são, geralmente, reversíveis, e podem coocorrer em uma mesma proteína, o que indica uma evidente modulação de cada tipo de modificação pós-traducional, como em uma espécie de código, onde cada combinação é lida de acordo com um padrão específico de PTMs, correspondendo a uma determinada função celular. Assim, pode-se dizer que os organismos dependem, em última instância, das proteínas em sua versão final, ou seja, modificada, para executar seus papéis, esse outro nível de informação confere a essas moléculas o aumento de sua complexidade (Figura 1.2).

Esse código foi inicialmente descrito em histonas (STRAHL; ALLIS, 2000), ficando conhecido como “código de histonas”, tem sido, principalmente na última década, ampliado para outros tipos de proteínas devido aos estudos de caracterização bioquímica e em larga escala por MS das PTMs (DAI; GU, 2010; ZHANG; WEN; SHI, 2012; BIGGAR; LI, 2014) .

Figura 1.2 - As PTMs aumentam o nível de complexidade nas células.



Fonte: adaptado de THERMO FISHER SCIENTIFIC (2016).

O DNA carrega a informação genética, cujos genes, que tem as instruções para a síntese de proteínas, são transcritos em moléculas de RNA mensageiros (mRNA), passam por processos de *splicing* alternativo, podendo gerar diferentes isoformas proteicas. Estes mRNA são traduzidos, tornam-se as proteínas e estas podem sofrer modificações, aumentando em complexidade o nível de informação. Assim, o proteoma de um organismo (o conjunto das proteínas deste) pode apresentar um conjunto de PTMs, como fosforilação, metilação, acetilação, poliubiquitinação, poli-ADP ribosilação, glicosilação, etc., estas modificações regulam a estrutura e as funções das proteínas, que em conjunto, desempenham inúmeras funções.

Com o propósito principal de gerenciar funções moleculares de maneira mais rápida, estas PTMs são extremamente dinâmicas em um organismo, permitindo uma rápida adaptação. Por exemplo, a adaptação a diferentes ambientes durante o ciclo de vida de um tripanossomatídeo, está intimamente relacionada com a expressão diferencial de genes e, conseqüentemente, à modulação de diferentes proteínas. Estas proteínas atingem suas formas maduras através destas modificações, pois elas determinam a especificidade de suas funções e auxiliam na rápida modulação dos diferentes processos celulares exigidos naquela determinada condição pela qual a célula está passando. As PTMs são uma das características moleculares mais relevantes no processo de sinalização celular, sendo cruciais às funções celulares e sobrevivência da célula.

Dentre mais de 200 PTMs conhecidas, as mais estudadas são a fosforilação, que ocorre em resíduos de serina, treonina e tirosina nas proteínas e está relacionada com a transdução de sinal, regulação da atividade enzimática, interação proteína-proteína e



proteínas-ligantes, dentre outros processos. A metilação, que ocorre em resíduos de lisina e arginina, está envolvida na interação proteína-proteína e proteína-ácido nucleico, regulação da cromatina e transcrição gênica (modificação de histonas), dentre outros processos. A glicosilação, que ocorre em resíduos de asparagina, serina e treonina, está envolvida com a estabilidade da proteína, solubilidade, sinal de secreção, regulação de interações, reconhecimento extracelular, etc. A acetilação, que ocorre em resíduos de lisina, está envolvida com a interação proteína-proteína e proteína-ácido nucleico e, como modificação em histonas, regula a transcrição gênica. A ubiquitinação, que ocorre em resíduos de lisina, é um importante sinal de degradação da proteína e também envolvida nas interações proteína-proteína (JENSEN, 2006).

### **1.3 Metilação de proteínas**

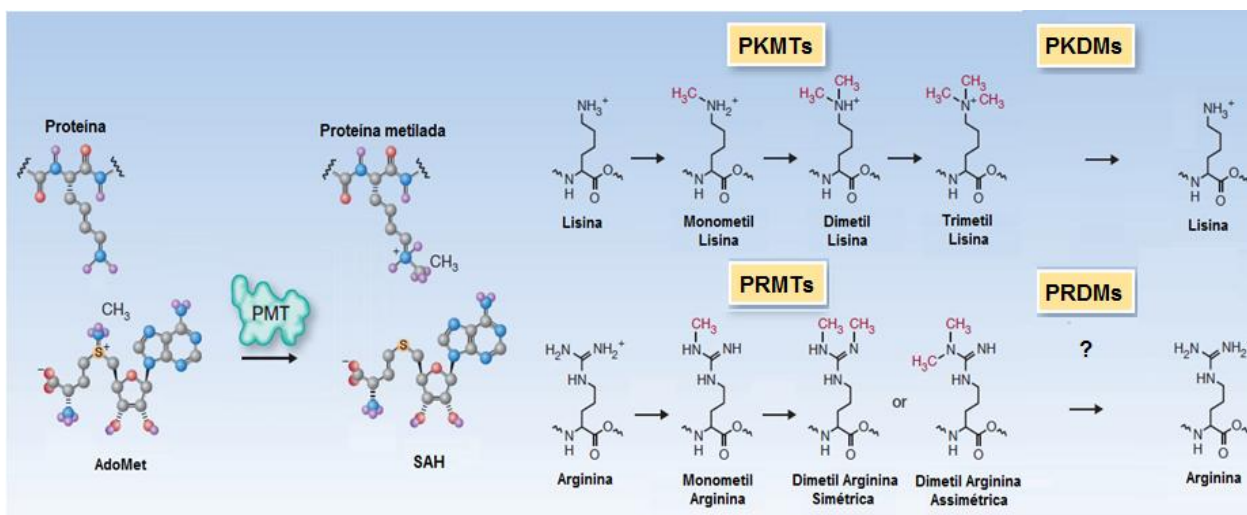
A metilação de proteínas é a catálise pela qual PMTs específicas, adicionam grupamentos metil ( $\text{CH}_3$ ) às proteínas. Essa PTM pode ocorrer em diferentes resíduos como histidina (H), cisteína (C), ácido aspártico (D), ácido glutâmico (E), serina (S) e treonina (T), no entanto, é mais comumente observada em resíduos de lisina (K) e arginina (R) (AFJEHI-SADAT; GARCIA, 2014). A metilação de lisinas e argininas ocorre quando uma PKMT e PRMT, respectivamente, utilizam AdoMet para adicionar grupos metil ao nitrogênio do grupo guanidina ( $\text{N}^G$ ) destes resíduos. O grupo  $\omega$ -amino dos resíduos de arginina podem ser monometilados (Me) ou dimetilados (Me<sub>2</sub>) por PRMTs e essa dimetilação pode ser simétrica, quando dois grupos metil são adicionados aos nitrogênios terminais adjacentes do grupo guanidino de um resíduo de arginina ou pode ser assimétrica, quando dois grupos metil são adicionados ao mesmo nitrogênio terminal. Nos resíduos de lisina, o grupo  $\epsilon$ -amino pode ser monometilado (Me), dimetilado (Me<sub>2</sub>) ou, ainda, trimetilado (Me<sub>3</sub>) por PKMTs, a dimetilação de lisina não apresenta a forma simétrica ou assimétrica como os resíduos de arginina, pois o grupo  $\epsilon$ -amino nos resíduos de lisina possui apenas um grupo amino terminal (Figura 1.3).

Assim como as fosfatases removem o fosfato dos substratos fosforilados, de modo análogo, as proteínas demetilases removem o grupo metil dos substratos que são

metilados. São conhecidas três classes de demetilases, as quais foram caracterizadas demetilando histonas (KLOSE; KALLIN; ZHANG, 2006). As demetilases da classe PADI4 (*Petidylarginine deiminase 4*) são conhecidas por remover o grupo metil em resíduos de arginina. Embora, seja claro seu papel na remoção do grupo metil, não é estritamente considerada uma demetilase, pois o produto da reação é uma citrulina e não um resíduo de arginina não-modificado. Em resíduos de lisina a metilação é notadamente reversível e ocorre através das proteínas lisina demetilases (PKDMs). A remoção do grupamento metil se dá por demetilases das classes LSD1 e Jumonji (Jmj) (CLOOS et al., 2008). Enzimas da classe LSD1 (*Lysine specific demethylase 1*) utilizam flavina como cofator e demetilam lisina mono ou dimetiladas (KLOSE; KALLIN; ZHANG, 2006). Uma terceira classe de enzimas demetilases são as que contêm o domínio Jumonji C (JmjC). A demetilação de lisina das histonas ocorre através de uma reação oxidante que requer ferro Fe (II) e  $\alpha$ -cetoglutarato ( $\alpha$ KG) como cofatores, podendo remover todos os três estados de metilação de lisina (Me, Me<sub>2</sub> e Me<sub>3</sub>) com representantes apresentando potencial atividade sobre proteínas não-histonas. Estas proteínas apresentam homologia entre domínios JmjC e cupin. Cupin é um domínio encontrado dentro de uma superfamília de proteínas que se caracterizam pela sua estrutura terciária de  $\beta$ -barril (KLOSE; KALLIN; ZHANG, 2006).

Em resíduos de arginina a reversibilidade é uma questão ainda aberta, uma vez que uma específica demetilase de arginina ainda não foi descrita, e estudos que identificaram possíveis candidatos (CHANG et al., 2007) mostraram resultados conflitantes posteriormente (BÖTTGER et al., 2015). Recentemente foi demonstrado que um subgrupo de demetilases JmjC tem duplo papel como demetilase tanto de lisina quanto de arginina, em histonas quanto não-histonas (WALPORT et al., 2016). No entanto, demetilases exclusivas de arginina ainda precisam ser identificadas.

Figura 1.3 - Catálise das PMTs a partir de AdoMet e os produtos da metilação de arginina e lisina



Fonte: adaptado de Copeland (2013).

S-adenosilmetionina (AdoMet) é o doador do grupo metil ( $\text{CH}_3$ ) na catálise pelas metiltransferases (PMTs), após a transferência torna-se S-adenosilhomocisteína (SAH) e a proteína é metilada no nitrogênio guanidino ( $\text{N}^G$ ). O produto desta reação em lisina (K) e arginina (R) está relacionado à quantidade de grupos metil ( $\text{CH}_3$ ) que o resíduo recebe das PKMTs e PRMTs, respectivamente. Lisinas e argininas podem ser mono e dimetiladas, sendo que em argininas essa dimetilação pode ocorrer de modo simétrico ou assimétrico. Lisinas podem ser trimetiladas. A demetilação de lisinas ocorre por demetilases das classes JmjC e LSD1. Por outro lado a demetilação de argininas por PRDMs ainda é pouco clara.

Enquanto que metilação de arginina tem sido relatada regulando o processamento de RNA, transcrição de genes, reparo de dano de DNA, translocação de proteína e transdução de sinal. A metilação de lisina é mais conhecida por regular a função das histonas e está envolvida na regulação epigenética da transcrição gênica (PAIK; PAIK; KIM, 2007; GUO et al., 2014). Além disso, outros fatores podem influenciar o impacto final da metilação, dependendo do sítio em que ela ocorre e *crosstalk* com outras modificações. A metilação já foi identificada em diferentes organismos, desde leveduras (WANG et al., 2014), tripanossomatídeos (LOTT et al., 2013; ROSENZWEIG et al., 2008) a células humanas (CAO; ARNAUDO; GARCIA, 2013; GUO et al., 2014; LARSEN et al., 2016), demonstrando sua importância em processos biológicos fundamentais na célula tendo em vista sua conservação em diferentes organismos eucariotos.

### 1.3.1 Metilação de arginina (R)

A metilação de argininas é uma modificação encontrada em células eucarióticas, realizada através de metiltransferases de arginina (PRMTs) responsáveis por mono e dimetilar o resíduo de argina das proteínas. As PRMTs são classificadas em quatro classes de acordo com o tipo de metilação que cada enzima realiza. O tipo mais comum é a classe do tipo I de enzimas, que catalisa a formação de arginina monometilada (Rme) no  $\omega$ -nitrogênio, seguido pela adição de um segundo grupo metil no mesmo  $\omega$ -nitrogênio, gerando assim uma dimetilarginina assimétrica (Rme2a). As PRMTs tipo II também catalisam a síntese de arginina monometilada  $\omega$ -nitrogênio, mas posteriormente adicionam um segundo grupo de metil, no nitrogênio terminal adjacente, resultando em dimetilarginina simétrica (Rme2s). A maioria das PRMTs é designada de tipo I ou tipo II. As PRMTs classe III catalisam a produção de apenas arginina monometilada (Me) no  $\omega$ -nitrogênio e foi relatada em *T. brucei*, assim como na levedura Hsl7, sendo que demonstra ter também atividade de tipo II. E o tipo IV de PRMT, catalisa Me no  $\delta$ -nitrogênio da arginina, isto é, o nitrogênio intermediário do resíduo, e foi até o momento demonstrado apenas em levedura (FISK et al., 2009). São conhecidas, até o momento, nove PRMTs em humanos (MORALES et al., 2016) e cinco PRMTs em *T. brucei* (FISK; READ, 2011).

Os substratos das PRMTs constituem uma fração cada vez maior do proteoma em diferentes organismos, conforme os estudos em larga escala vêm sendo realizados. Estas proteínas estão comumente envolvidas comm transdução de sinal, *splicing* de RNAm, controle de transcrição, reparo de DNA e translocação de proteína (BEDFORD; CLARKE, 2009).

### 1.3.2 Metilação de lisina (K)

A metilação de lisina foi pela primeira vez reportada por Ambler e Rees (1959), na proteína flagelina da *Salmonella typhimurium*, e foi verificado pouco tempo depois em histonas por Murray (1964), e desde então cada vez mais proteínas tem sido demonstradas sendo impactadas por metilação de lisinas e que esta é uma modificação

de ampla ocorrência nas células tanto procarióticas quanto eucarióticas.

As PKMTs utilizam o mesmo precursor AdoMet que as PRMTs, e catalisam a adição de até três grupos metil ao  $\epsilon$ -amino do resíduo de lisina. As PKMTs agrupam as enzimas que contêm um domínio catalítico SET, em homenagem a SU (var), Enhancer de Zeste e Trithorax, as três primeiras proteínas identificadas portadoras deste domínio em *Drosophila* (TSCHIERSCH et al., 1994). Com base na sequência de aminoácidos primária, nos domínios e especificidade de substrato, as PKMTs são classificadas em oito subfamílias distintas (PKMT1 a PKMT8). Com exceção de um segundo grupo de PKMTs que possuem o domínio DOT (*Disruptor of Telomeric Silencing*) e metilam histonas, todas as demais PKMTs conhecidas contêm o domínio SET e podem metilar tanto histonas quanto proteínas não-histonas (ZHANG; WEN; SHI, 2012; ZHANG et al., 2015; FAROOQ et al., 2016).

A metilação de lisinas foi inicialmente demonstrada atuando, sobretudo, em processos relacionados à epigenética. Com o avanço de estratégias proteômicas, tem sido reportada influenciando a função das proteínas além do contexto da cromatina, modulando a deposição de outras PTMs, tais como fosforilação, acetilação e ubiquitinação ou regulando as interações proteína-proteína. Por exemplo, em eucariotos a metilação de lisina está presente em chaperonas como HSP90 e HSP70, promovendo a associação com outras proteínas; está presente em Dam1 em leveduras regulando a segregação dos cromossomos durante a divisão celular. Em procariotos, a metilação de lisina regula a função de proteínas envolvidas com motilidade ou então em proteínas de superfície atua otimizando a adesão bacteriana ao seu ambiente. Além disso, alguns vírus, como o HIV-1 utilizam as PKMTs de seus hospedeiros para metilar diferentes sítios em suas próprias proteínas, como o ativador transcricional Tat, desse modo ativando ou reprimindo a transcrição de suas proteínas (LANOUILLE et al., 2014).

#### **1.4 Metilproteoma de tripanossomatídeos**

A publicação dos genomas dos tripanossomatídeos *T. cruzi*, *T. brucei* e *Leishmania major* (*TriTryps*) em 2005 (BERRIMAN et al., 2005; EL-SAYED et al., 2005b; IVENS et al., 2005) facilitou os estudos ômicos, principalmente os de “larga

escala”. Percebeu-se que estes tripanossomatídeos compartilham diversas características celulares e genéticas, como funções de organelas e características do kDNA, de modo que estas semelhanças auxiliam no estudo comparativo destes organismos. O alinhamento da sequência aminoacídica entre *T. cruzi* e *T. brucei* revelou aproximadamente 57% de identidade (EL-SAYED et al., 2005a). Embora a maioria dos genes nestes organismos tenha se revelado conservada, há diferenças substanciais que podem ser exploradas em nível de proteína através da MS, principalmente no que concerne às PTMs.

O estudo das PTMs em tripanossomatídeos passou a ser possível em “larga escala” em associação com a proteômica associada à MS e tem sido demonstrada principalmente em *Leishmania donovani* (ROSENZWEIG et al., 2008) e *T. brucei* (FISK et al., 2013; LOTT et al., 2013) e em *T. cruzi* (DE JESUS et al., 2016). Em *L. donovani*, o estudo identificou metilação, fosforilação, acetilação e glicosilação durante a diferenciação deste organismo em diferentes proteínas, já em *T. cruzi* foram identificadas as PTMs presentes nas histonas das formas epimastigota e tripomastigota deste organismo.

Em *T. brucei*, estes dois recentes trabalhos apresentaram resultados expressivos quanto o impacto das metilações em arginina neste parasito, de modo que os seus dados sobre o metilproteoma de arginina em *T. brucei* permitiram estimar que aproximadamente 10% do proteoma deste parasito contém metilargininas, demonstrando ser uma modificação abrangente, que afeta processos como a regulação gênica, reparo de DNA, tráfego de proteínas, sinalização, *foldi*ng, degradação proteica, além de atuar nas funções flagelares (LOTT et al., 2013). A presença de metilação em proteínas em tripanossomatídeos destaca seu potencial de impactar processos cruciais a estes organismos. Seguindo os estudos dentro do *TriTryp*, *T. cruzi* é provavelmente o que carece de estudos sobre o impacto das PTMs em sua biologia, são escassos e muito recentes os trabalhos neste organismo.

## 1.5 Estudo das PTMs por MS

As PTMs são conhecidas há décadas, no entanto, a compreensão de o quão abrangente elas são começou a ser revelado apenas ao longo da última década através de MS, pois permitiu mapear essas modificações em organelas ou mesmo células inteiras e identificar suas mudanças quantitativas nas células em diferentes condições (OLSEN; MANN, 2013). A aplicação de MS na identificação de modificações em proteínas purificadas não é recente (STENFLO et al., 1974), mas o mapeamento em larga escala de modificações tem sido, assim como o emprego de abordagens específicas na análise de cada PTM. O estudo das proteínas e suas modificações por MS constituem o próximo nível nos esforços para compreensão da expressão gênica (OLSEN; MANN, 2013).

Algumas peculiaridades existem neste tipo de análise. Enquanto que na análise de proteínas muitos peptídeos estão disponíveis para a posterior identificação de cada proteína, na análise das PTMs cada peptídeo com um sítio modificado deve ser identificado e validado individualmente, e nem sempre estes peptídeos estão presentes de maneira abundante em relação às suas versões não modificadas (OLSEN; MANN, 2013).

De modo geral, os métodos de preparação de amostra são muito similares aos utilizados em uma análise proteômica *shotgun* padrão, onde o conjunto de proteínas de uma organela ou célula total é digerido por uma protease específica, geralmente tripsina, e os peptídeos são separados por cromatografia, geralmente por fase reversa e analisados por cromatografia líquida acoplada a espectrometria de massas sequencial (LC-MS/MS). A maior exceção está na etapa de enriquecimento da modificação específica, fundamental para a sua ampla identificação, mas com estratégias disponíveis para poucas PTMs. Por exemplo, na identificação de fosforilação se utiliza, geralmente, cromatografia baseada na afinidade por dióxido de titânio (TiO<sub>2</sub>) ou anticorpos antifosfotirosina; para metilação pode-se utilizar anticorpos específicos para mono e dimetilação de arginina; para glicopeptídeos N-ligados são utilizadas lectinas; e anticorpos específicos para di-glicina para detectar sítios de ubiquitinação. No entanto, mesmo com estes enriquecimentos a identificação pode

encontrar algumas dificuldades, como quando estas modificações estão em proteínas transmembrana (OLSEN; MANN, 2013). Além disso, estratégias de fragmentação durante a análise por MS podem auxiliar muito a identificação dos sítios modificados, por exemplo, como *multistage activation* (MSA) (SCHROEDER et al., 2004) e *higher collisional dissociation* (HCD) (OLSEN et al., 2007).

## 1.6 Estudo do metilproteoma por MS

A metilação proteica tem sido estudada, de modo tradicional, utilizando abordagens, tais como anticorpos específicos de metilação, mapeamento de PTMs por MS e marcação radioativa para caracterizar a metilação das proteínas alvo. As abordagens mais recentes têm proporcionado novas maneiras para identificar proteínas metiladas, medir os níveis de metilação, identificar substratos de enzimas metiltransferases e combinar proteínas metiladas a domínios específicos de metilação. Domínios proteicos de ligação a metilação, como o 3xMBT descrito por Moore e colaboradores (2013) e anticorpos aprimorados com ampla especificidade para proteínas metiladas (CAO; ARNAUDO; GARCIA, 2013; GUO et al., 2014) estão sendo utilizados para caracterizar o metilproteoma. Em outras abordagens, a marcação deixa de ser radioativa e passa a utilizar marcação metabólica com aminoácidos estáveis, empregando SILAC (ONG; MITTLER; MANN, 2004).

O constante aprimoramento das estratégias utilizadas no estudo de proteínas metiladas, objetivo este que também faz parte do escopo deste trabalho, tem buscado aumentar a eficiência na identificação de novos sítios e proteínas metiladas e ao mesmo tempo otimizar as etapas experimentais, como por exemplo, algumas estratégias incluem um único passo de purificação por afinidade dos peptídeos metilados, não utilizando pré-fracionamento da amostra, como demonstrado no protocolo de Uhlmann e colaboradores (2012) e Bremang e colaboradores (2013). Por outro lado, há estratégias que pré-fracionam a amostra utilizando SCX (*Strong Cation Exchange*) e enriquecem a nível peptídico a amostra com anticorpos comerciais específicos, como demonstrado por Sylvestersen e Nielsen (2015).



Estas ferramentas são, muitas vezes, acopladas tanto a melhorias em MS, a fim de identificar rapidamente os resíduos metilados quanto em bioinformática, com ferramentas de busca e análise dos dados proteômicos. A maquinaria envolvida na modificação também é alvo e constantes novas técnicas bioquímicas empregadas para sondar a função de metiltransferases, demetilases, e os domínios leitores de ligação de metil. Associado a isso estão ferramentas de bioinformática que permitam a analisar os dados gerados. Exemplo disso é o agrupamento em hierarquias das anotações gênicas das proteínas identificadas através das ontologias gênicas. Por exemplo, as ontologias gênicas (*Gene Ontology*, GO) referem-se ao produto de um determinado gene e à função e localização que ele desempenha na maquinaria celular. São classificadas em três níveis hierárquicos onde agrupam-se as proteínas identificadas com base na sua ontologia gênica: (i) Componente celular ou localização celular refere-se ao local da célula onde a proteína está ativa; (ii) Processo biológico, refere-se ao objetivo biológico ao qual a proteína contribui e (iii) Função molecular, descreve a atividade bioquímica que uma dada proteína desempenha no meio celular (ASHBURNER et al., 2000).

Estas ferramentas possibilitam um entendimento sistêmico de um proteoma, e de seus subproteomas, como o metilproteoma, integrando as proteínas em processos de sinalização mais amplos, além de potencializarem sua utilização em análises e ensaios em larga escala na busca por inibidores e desenvolvimento de drogas (CARLSON; GOZANI, 2014).

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Caracterizar o metilproteoma de formas epimastigotas de *Trypanosoma cruzi*.

### **2.2 Objetivos específicos**

- Analisar, *in silico*, a presença da maquinaria de metilação e demetilação de *T. cruzi* e sua similaridade com *T. brucei* e humanos;
- Otimizar o preparo de amostras de *T. cruzi* para identificação de sítios de metilação por LC-MS/MS;
- Identificar, em larga escala, proteínas alvo e respectivos sítios de metilação da forma epimastigota de *T. cruzi*;
- Desenhar um panorama geral das vias reguladas por metilação em formas epimastigotas de *T. cruzi*.

### 3 METODOLOGIA

#### 3.1 Busca por maquinaria de metilação/demetilação, domínios conservados e alinhamentos de sequência primária

Para a busca das PRMTs de *T. cruzi*, foram utilizadas duas estratégias: (i) o algoritmo Protein BLAST (*Basic Alignment Search Tool*, <https://blast.ncbi.nlm.nih.gov/>), disponível no NCBI (*National Center of Biotechnology Information*, <http://www.ncbi.nlm.nih.gov>), onde as sequências das PRMTs de *T. brucei* TREU927 e *Homo sapiens* (FISK; READ, 2011) foram utilizadas como *query* na identificação das sequências proteicas mais similares (sequências com maior *score* e menor *e-value*) em *T. cruzi* CL Brener e (ii) a ferramenta *gene text search* da plataforma TriTrypDB (<http://tritrypdb.org/tritrypdb>), buscando o termo “*arginine N-methyltransferase*” para recuperação das proteínas anotadas em *T. cruzi* CL Brener. Para a busca das PKMTs descritas em *T. cruzi* e *T. brucei* foi utilizada a ferramenta *gene text search* da plataforma TriTrypDB e em *Homo sapiens* foi buscado no GenBank (<http://www.ncbi.nlm.nih.gov/genbank>), buscando o termo “*lysine N-methyltransferase*” e a correlação entre as homólogas foi confirmada por BLAST. Para a busca das demetilases foram resgatadas as proteínas com domínio Jmjc (Interpro IPR003347) na plataforma TriTrypDB e as homólogas identificadas por BLAST.

A busca por domínios conservados nas sequências proteicas foi realizada através da plataforma CDD (*Conserved Domains Database*; <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), considerando *e-value* de 0.01 (MARCHLER-BAUER et al., 2015).

O Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo>) dentro da base de dados Uniprot foi empregado no alinhamento e análise de similaridade das sequências das PRMTs, PKMTs e demetilases e a edição das imagens foi feita através do Jalview (<http://www.jalview.org>).

### 3.2 Cultura de células

Células epimastigotas de *T. cruzi* Dm28c em fase de crescimento exponencial foram mantidas em cultura axênica em meio LIT (*Liver Infusion Tryptose*) suplementado com 10% de soro bovino fetal a 28 °C, com passagens a cada 3-4 dias. Foram obtidas células de epimastigota na ordem de  $4,3 \times 10^9$ . Este total de células foi dividido em cinco tubos, cada um com aproximadamente  $8,6 \times 10^8$  células. Estas células de cada tubo foram centrifugadas a 5000 *xg*, por 10 minutos a 4°C, lavadas em 5 mL de PBS 1X, centrifugadas novamente a 5000 *xg*, por 10 minutos a 4 °C, ressuspensas em 1 mL de PBS 1X, centrifugadas novamente nas mesmas condições anteriores e os *pellets* foram congelados (-20°C) até o uso.

### 3.3 Extração e digestão de proteínas em solução

Uma vez obtidas as células de epimastigotas, foi realizada a extração das proteínas utilizando a proporção de 240 µL de tampão de lise (SDS 4%, Tris-HCl 10 mM pH 7.5, DTT 100 mM) para cada  $3 \times 10^8$  células. Os *pellets* foram ressuspensos no tampão de lise através de micropipetagem, as amostras agitadas em vortex por 30 segundos, aquecidas por 3 minutos a 95 °C e em seguida sonicadas por 1 hora, até não apresentarem mais viscosidade. Então, os extratos foram centrifugados a 20.000 *xg* por cinco minutos, a 20 °C. Em seguida, as amostras foram armazenadas a -80 °C.

A digestão das proteínas foi realizada utilizando *Filter-Aided Sample Preparation* (FASP) (WIŚNIEWSKI et al., 2009), onde em uma coluna de Microcon YM 30K (Millipore) foi misturada uma quantidade de amostra equivalente a 200 µg de proteínas com 7 vezes o volume de solução A (Ureia 8 M, Tris-HCl 100 mM pH 8.8). A amostra foi homogeneizada por micropipetagem, centrifugada a 14.000 *xg* por 15 minutos a 20 °C e o *flowthrough* foi descartado. Na sequência, foram adicionados 200 µL de solução B (DTT 10 mM em A) à coluna e misturado com micropipeta, centrifugado a 14.000 *xg*, por 15 minutos, a 20 °C. Descartado o *flowthrough*. Esta etapa foi repetida por mais duas vezes, sendo que a cada nova lavagem o lado da membrana do Microcon que fica voltado para o interior do rotor da centrífuga foi alternado. Após, foram adicionados 100

$\mu\text{L}$  de solução C (Iodoacetamida 50 mM em A), misturado com micropipeta, homogeneizado por 1 minuto, a 600 rpm, 20 °C em thermomixer, e então, incubada a amostra por 20 minutos no escuro. Na sequência, centrifugado a 14.000  $xg$ , por 10 minutos, a 20 °C. Logo após, foram feitas duas etapas de lavagem com 100  $\mu\text{L}$  de solução A, seguida de duas etapas de lavagem com 100  $\mu\text{L}$  de bicarbonato de amônio ABC (50 mM), sendo que, em cada lavagem, a amostra foi homogeneizada com micropipeta e centrifugada a 14.000  $xg$ , por 15 minutos, a 20 °C. A cada lavagem, o lado da membrana do Microcon voltada para o interior do rotor da centrífuga foi alternado.

A amostra foi então dosada por Qubit 2.0 Fluorometer (Life Technologies). Na sequência, foi adicionada tripsina (Sequence Grade Modified, Promega) na proporção de 1  $\mu\text{g}$  de tripsina: 100  $\mu\text{g}$  de proteína, completado com ABC até o volume de 40  $\mu\text{L}$ , e incubado no Microcon em câmara úmida por 18 horas a 37 °C. Após a digestão, a amostra foi centrifugada em um novo tubo Microcon, a 14.000  $xg$ , por 10 minutos, a 20 °C, para coletar os peptídeos. Foram feitas então duas etapas de eluição, uma com 40  $\mu\text{L}$  de ABC (50 mM) e outra com 50  $\mu\text{L}$  de NaCl (0,5 M), homogeneizando com micropipeta e centrifugando a 14.000  $xg$ , por 15 minutos, a 20 °C. Os peptídeos foram então dosados no NanoDrop ND 1000 (Thermo Scientific) e a amostra acidificada até uma concentração final de 0,5% de TFA.

### **3.4 Fracionamentos de peptídeos e proteínas**

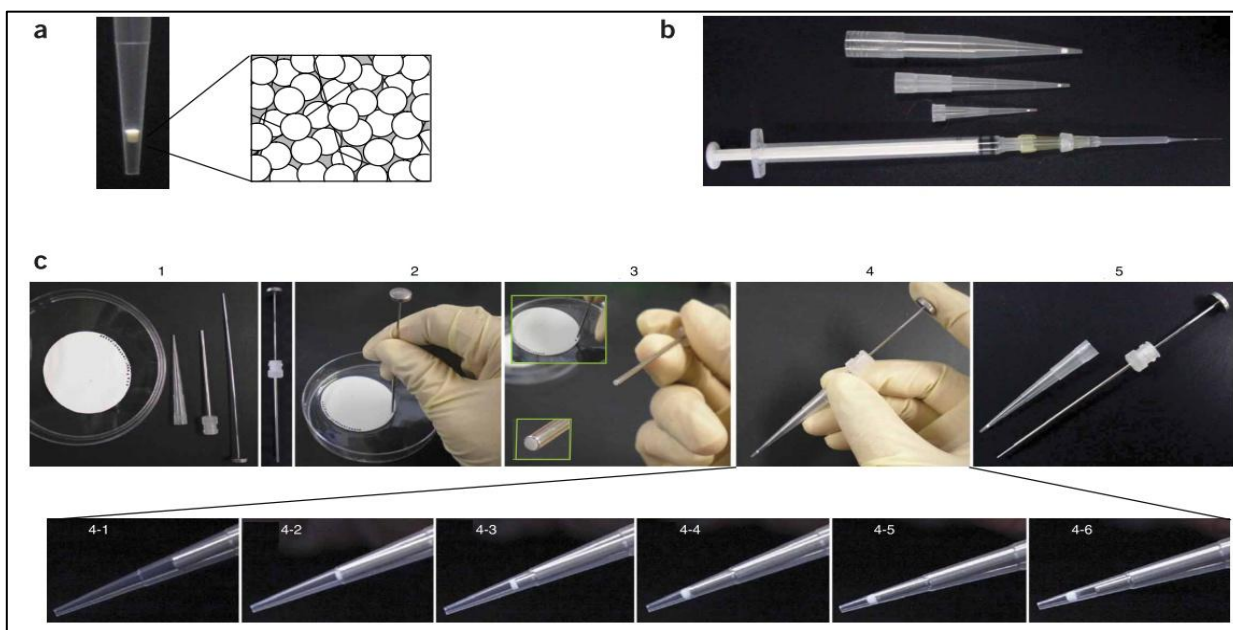
#### **3.4.1 Produção das *StageTips* (*Stop-and-Go-extraction Tips*)**

Os fracionamentos de peptídeos foram realizados empregando *StageTips* (*Stop-and-Go-extraction Tips*) (RAPPSILBER; ISHIHAMA; MANN, 2003) confeccionadas *in-house* a partir da colocação de discos 3M Empore™ que possuem *beads* cromatográficas (C18 e/ou SCX), imobilizadas em uma malha de Teflon (Figura 3.1a) em ponteiros (*tips*) de 200  $\mu\text{L}$ . Pode-se utilizar diferentes tamanhos de ponteira de acordo com a capacidade de carregamento desejada (Figura 3.1b).

Os discos estavam previamente armazenados em uma placa de Petri e utilizou-

se uma pinça previamente limpa para manusear o disco (Figura 3.1c 1). Em seguida, foi pressionado, gentilmente, o cortador (seringa cujo êmbolo interno é uma haste oca preenchida por um pistão) no disco (Figura 3.1c 2) e certificado que o material preencheria o interior de todo o êmbolo (Figura 3.1c 3). Colocado o cortador dentro da ponteira e soltado o disco empurrando o pistão até liberar a porção do disco dentro da mesma (Figura 3.1c 4), pressionado o disco, gentilmente, até a ponta da ponteira (Figura 3.1c 4-1 a 4-6). Enfim, removido o cortador (Figura 3.1c 5) (RAPPSILBER; MANN; ISHIHAMA, 2007). Foi utilizada a proporção de uma porção de cada disco para cada 5  $\mu\text{g}$  de peptídeo (RAPPSILBER; ISHIHAMA; MANN, 2003; WIŚNIEWSKI; ZOUGMAN; MANN, 2009).

Figura 3.1 - Produção de uma *StageTip* padrão



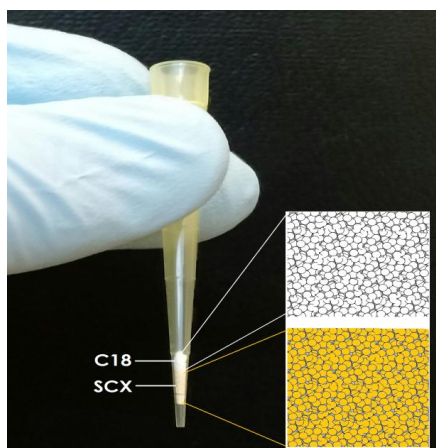
Fonte: adaptado de Rappsilber, Mann e Ishihama (2007).

### 3.4.2 Fracionamento por cromatografia de troca iônica - C18-SCX (sal)

Neste protocolo foram empregadas *StageTips* utilizando discos C18 e SCX (3M Empore™) combinados em uma mesma *StageTip* (RAPPSILBER; MANN; ISHIHAMA, 2007). Uma adaptação feita a esse protocolo foi a montagem de *StageTips* com cinco

porções de cada tipo de disco, para utilização de maior quantidade de amostra por *StageTip* (Figura 3.2). Uma vez que neste protocolo as frações são obtidas a partir dos tampões com diferentes concentrações de  $\text{NH}_4\text{AcO}$ , este protocolo foi denominado C18-SCX (**sal**).

Figura 3.2 - *StageTip* C18-SCX com 5 porções de cada disco



Fonte: o autor.

A porção C18 (em branco) corresponde às porções de disco que possuem as *beads* com propriedades cromatográficas de fase reversa. Já SCX (em amarelo), representa a região da *StageTip* com as porções de disco que possuem as *beads* com propriedades cromatográficas de troca catiônica.

Os tampões utilizados nesse fracionamento foram denominados **ReX-Buffer** (*Reversed phase-ion exchange*) pois o fracionamento apresenta duas propriedades de fracionamento combinadas, fase reversa e troca iônica. Estes tampões possuem as seguintes concentrações de metanol, acetato de amônio ( $\text{NH}_4\text{AcO}$ ), ácido acético (AcOH) e acetonitrila (ACN):

- Metanol 100%
- ReX-Buffer 1 - 0.5% AcOH
- ReX-Buffer 2 - 0.5% AcOH, 80% ACN
- ReX-Buffer 3 - 20 mM  $\text{NH}_4\text{AcO}$ , 0.5% AcOH, 20% ACN
- ReX-Buffer 4 - 50 mM  $\text{NH}_4\text{AcO}$ , 0.5% AcOH, 20% ACN
- ReX-Buffer 5 - 100 mM  $\text{NH}_4\text{AcO}$ , 0.5% AcOH, 20% ACN
- ReX-Buffer 6 - 500 mM  $\text{NH}_4\text{AcO}$ , 0.5% AcOH, 20% ACN

Após a digestão ter sido feita por FASP foi utilizada uma quantidade de 50 µg de peptídeos para este fracionamento. Para essa quantidade de amostra foram montadas 2 *StageTips* C18-SCX com 5 porções de cada.

Um procedimento fundamental para a adequada utilização das *StageTips* foi realizar seu condicionamento antes de carregar a amostra. Para condicionar cada porção de disco, tanto C18 quanto SCX, foram usados 20 µL de cada reagente, assim proporcionalmente cada *StageTip* recebeu 200 µL de metanol 100% seguido da centrifugação a 3300 rpm, 200 µL de ReX-Buffer 2 seguido da centrifugação a 3300 rpm, 200 µL de ReX-Buffer 1, seguido da centrifugação a 3300 rpm, 200 µL de ReX-Buffer 6, centrifugando a 3300 rpm (o condicionamento com este ReX-Buffer é fundamental para este tipo de fracionamento) e 200 µL de ReX-Buffer 1, centrifugando a 3300 rpm.

Foi tomado o cuidado para não deixar secar a *StageTip* em nenhuma etapa do condicionamento, principalmente na etapa final com ReX-Buffer 1, pois isso comprometeria seu condicionamento e assim a ligação dos peptídeos à resina. Após as *StageTips* estarem condicionadas, a amostra foi diluída em ReX-Buffer 1 (1:1).

Foram montadas duas *StageTips*, onde cada uma foi carregada com 25 µg da amostra, ou seja, foi considerada a capacidade de que cada porção de disco, nas dimensões que possui, pode carregar até 5 µg, por isso as 5 porções de C18 e 5 de SCX em cada tip. Adicionados 100 µL da amostra diluída em cada uma das *StageTip* e centrifugado por 3 minutos a 3300 rpm coletando o *flowthrough* e passando novamente a amostra nas *StageTips*. Foi coletado o *flowthrough* C18 (chamado FT C18 para posterior análise). Em seguida, foram lavadas as *StageTips* com 100 µL de ReX-Buffer 1, eluídas com 100 µL do ReX-Buffer 2. Novamente coletado o *flowthrough* SCX (FT SCX).

Na sequência, a fração 1 foi obtida eluindo as *StageTips* com 100 µL do ReX-Buffer 3; a fração 2 foi obtida eluindo as *StageTips* com 100 µL do ReX-Buffer 4; a fração 3 foi obtida eluindo as *StageTips* com 100 µL do ReX-Buffer 5; a fração 4 foi obtida eluindo as *StageTips* com 100 µL do ReX-Buffer 6.

As frações e FTs foram então quantificadas por NanoDrop ND 1000 (Thermo Scientific) e, em seguida, carregados em *StageTips* com porções de disco C18. Cada



*StageTip* C18 foi montada com uma quantidade de porções correspondente com a quantidade de amostra da fração/FT que recebeu, considerando a capacidade supracitada de 5 ug por porção de disco. Para condicionar cada *StageTip* C18, foi mantida a proporção de 20 µL de cada um dos reagentes (Metanol 100%, ReX-Buffer 2 e ReX-Buffer 1, nessa sequência) por porção de disco, centrifugando sempre a 3300 rpm.

Então, foi realizado o carregamento da amostra nas *StageTips* C18. Para isso, foram adicionados 300 µL de Rex-Buffer 1 a cada fração/FT, misturado e passado em sua respectiva *StageTip* C18. Cada amostra foi passada duas vezes. As *StageTips* C18 foram lavadas com 300 µL de ReX-Buffer 1, centrifugando a 3300 rpm, por aproximadamente 2 minutos. As *StageTips* C18 foram então armazenadas a -20 °C até a análise por LC-MS/MS.

### 3.4.3 Fracionamento por cromatografia de troca iônica - C18-SCX (pH)

Uma vez que neste protocolo as frações são obtidas por cromatografia em *StageTip* multifuncional que inclui troca iônica, a partir dos tampões com diferentes valores de pH (RAPPSILBER; MANN; ISHIHAMA, 2007), este protocolo foi denominado C18-SCX (pH).

Para o fracionamento por C18-SCX (pH), as *StageTips* foram montadas como descrito no item 3.4.1, os tampões utilizados foram denominados **ReX-Buffer** com as seguintes concentrações de metanol, NH<sub>4</sub>AcO, bicarbonato de amônio (NH<sub>4</sub>HCO<sub>3</sub>), hidróxido de amônio (NH<sub>4</sub>OH), ácido acético (AcOH) e acetonitrila (ACN):

- Metanol 100%
- ReX-Buffer 1 - 0.5% AcOH
- ReX-Buffer 2 - 0.5% AcOH, 80% ACN
- ReX-Buffer 3 - 50 mM NH<sub>4</sub>AcO, 20% ACN
- ReX-Buffer 4 – 50 mM NH<sub>4</sub>HCO<sub>3</sub>, 20% ACN
- ReX-Buffer 5 - 0.1% NH<sub>4</sub>OH, 20% ACN
- ReX-Buffer 6 - 500 mM NH<sub>4</sub>AcO, 0.5% AcOH, 20% ACN

Após FASP, foram preparadas 3 *StageTips* C18-SCX com 5 porções de disco C18 e 5 porções de disco SCX, pois foram fracionados 75 µg de peptídeos, logo cada

uma recebeu 25 µg de peptídeos após o condicionamento.

As *StageTips* C18-SCX foram condicionadas como descreve o item 3.4.2, inclusive com a etapa que utiliza o ReX-Buffer 6, pois este condiciona o material de troca iônica (SCX) para a adequada ligação dos peptídeos.

Após cada *StageTips* ser carregada com 25 µg de peptídeos, foram lavadas com 100 µL do pH-Buffer 1 e fracionadas eluindo as *StageTips* com 100 µL do ReX-Buffer 3, obtendo a fração 1; eluindo as *StageTips* com 100 µL do ReX-Buffer 4, obtendo a fração 2; eluindo as *StageTips* com 100 µL do pH-Buffer 5, obtendo a fração 3.

As frações foram então quantificadas por NanoDrop ND 1000 (Thermo Scientific) e, em seguida, carregadas em *StageTips* com porções de disco C18. Previamente, cada *StageTip* C18 foi montada com uma quantidade de porções correspondente com a quantidade de amostra da fração que recebeu, considerando a capacidade de 5 µg por porção de disco. Para condicionar cada *StageTip* C18, foi mantida a proporção de 20 µL de cada um dos reagentes (Metanol 100%, ReX-Buffer 2 e ReX-Buffer 1, nessa sequência) por porção de disco, centrifugando sempre a 3300 rpm.

Então, foi realizado o carregamento da amostra nas *StageTips* C18. Para isso, foram adicionados 300 µL de ReX-Buffer 1 a cada fração, misturado e passado em sua respectiva *StageTip* C18. Cada amostra foi passada duas vezes. As *StageTips* C18 foram lavadas com 300 µL de ReX-Buffer 1, centrifugando a 3300 rpm, por aproximadamente 2 minutos. As *StageTips* C18 foram então armazenadas a -20 °C até a análise por LC-MS/MS.

#### 3.4.4 Fracionamento por cromatografia de fase reversa - C18 (pH ácido)

Para o fracionamento por C18 (pH ácido), as *StageTips* foram montadas como descrito no item 3.4.1, mas apenas com as porções de C18 da Figura 3.2.

Os tampões utilizados nesse fracionamento foram denominados **pH-Buffer** e possuem as seguintes concentrações de metanol, ácido acético (AcOH) e acetonitrila (ACN) (RAPPSILBER; MANN; ISHIHAMA, 2007):

- Metanol 100%
- pH-Buffer 1 - 0.5% AcOH
- pH-Buffer 2 - 0.5% AcOH, 10% ACN
- pH-Buffer 3 - 0.5% AcOH, 20% ACN
- pH-Buffer 4 - 0.5%,AcOH, 80% ACN

Uma vez que neste protocolo as frações são obtidas a partir de cromatografia reversa a diferentes valores de pH este protocolo foi denominado C18 (**pH ácido**).

Após FASP, foram preparadas 3 *StageTips* com 5 porções de disco C18, pois foram fracionados 75 µg de peptídeos, logo cada uma recebeu 25 µg de peptídeos após o condicionamento.

Para condicionar as *StageTips* C18, foram usados 20 µL de cada reagente, assim proporcionalmente cada *StageTip* recebeu 100 µL de metanol 100 % seguido da centrifugação a 3300 rpm, 100 µL de pH-Buffer 4 seguido da centrifugação a 3300 rpm, 100 µL de pH-Buffer 1, seguido da centrifugação a 3300 rpm.

Após cada *StageTips* ser carregada com 25 µg de peptídeos, foram lavadas com 100 uL do pH-Buffer 1 e fracionadas eluindo as *StageTips* com 100 µL do pH-Buffer 2, obtendo a fração 1; eluindo as *StageTips* com 100 µL do pH-Buffer 3, obtendo a fração 2; eluindo as *StageTips* com 100 µL do pH-Buffer 4, obtendo a fração 3.

As frações foram então quantificadas por NanoDrop ND 1000 (Thermo Scientific) e, em seguida, carregadas em *StageTips* com porções de disco C18. Previamente, cada *StageTip* C18 foi montada com uma quantidade de porções correspondente com a quantidade de amostra da fração que recebeu, considerando a capacidade de 5 ug por porção de disco. O condicionamento de cada *StageTip* C18 nesta etapa se deu como descrito acima neste mesmo protocolo.

Então, foi realizado o carregamento da amostra nas *StageTips* C18. Para isso, foram adicionados 300 µL de pH-Buffer 1 a cada fração, misturado e passado em sua respectiva *StageTip* C18. Cada amostra foi passada duas vezes. As *StageTips* C18 foram então lavadas com 300 µL de ReX-Buffer 1, centrifugando a 3300 rpm, por aproximadamente 2 minutos. As *StageTips* C18 foram então armazenadas a -20 °C até a análise por LC-MS/MS.

### 3.4.5 Fracionamento por cromatografia de fase reversa - C18 (pH básico)

Neste protocolo foi utilizada uma coluna Sep Pak C18 (130 mg) para o carregamento e fracionamento dos peptídeos, de acordo com (LUCENA, 2015). Os reagentes utilizados foram metanol 100% e formiato de amônio (FA) 200 mM (hidróxido de Amônio 0,8%, ácido fórmico 0,16%), pH 10, que foi posteriormente diluído até 20 mM, Acetonitrila (ACN). As etapas de equilíbrio, carregamento, lavagem e eluição, foram feitas utilizando uma seringa (5 mL) acoplada à coluna Sep Pak C18.

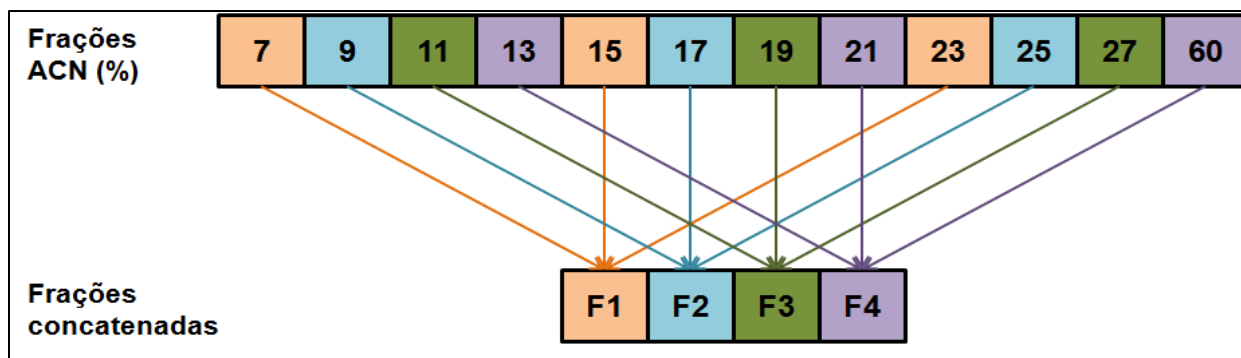
A coluna foi ativada com 5 mL metanol 100% e equilibrada com 5 mL de FA 20 mM (pH 10). Foram obtidos 340 ug de peptídeos por FASP, e sua amostra teve seu volume completado para 10 mL de FA 20 mM. A amostra foi então carregada na coluna por duas vezes e em seguida foi lavada por duas vezes com 5 mL de FA 20 mM.

Na sequência foram feitas 2 eluições com 500 uL com 12 diferentes soluções de FA 20 mM com ACN variando de 7 a 60% (Figura 3.3). Os eluatos foram dosados em NanoDrop ND 1000 (Thermo Scientific). As frações foram concentradas em Speed Vac (Thermo Scientific), pelo tempo necessário para evaporar 95% da ACN das amostras.

Em seguida, as amostras foram eluídas em 40 uL de Buffer A1 (0,1% ácido fórmico), dessalinizadas em *StageTips* C18, centrifugando a 3300 rpm, lavadas 3 vezes com Buffer A, e eluídas das *StageTips* C18 com 20 uL de Buffer B (40% ACN e 0,1% ácido fórmico), secadas em Speed Vac e eluídas com 10 uL de Buffer A2 (0,1% de ácido fórmico, 5% de DMSO).

Estes eluatos foram combinados de modo intercalado (Figura 3.3), gerando quatro frações maiores que foram analisadas por LC-MS/MS (YANG et al., 2012)

Figura 3.3 - Estratégia de concatenação das frações aplicada neste fracionamento.



Fonte: o autor.

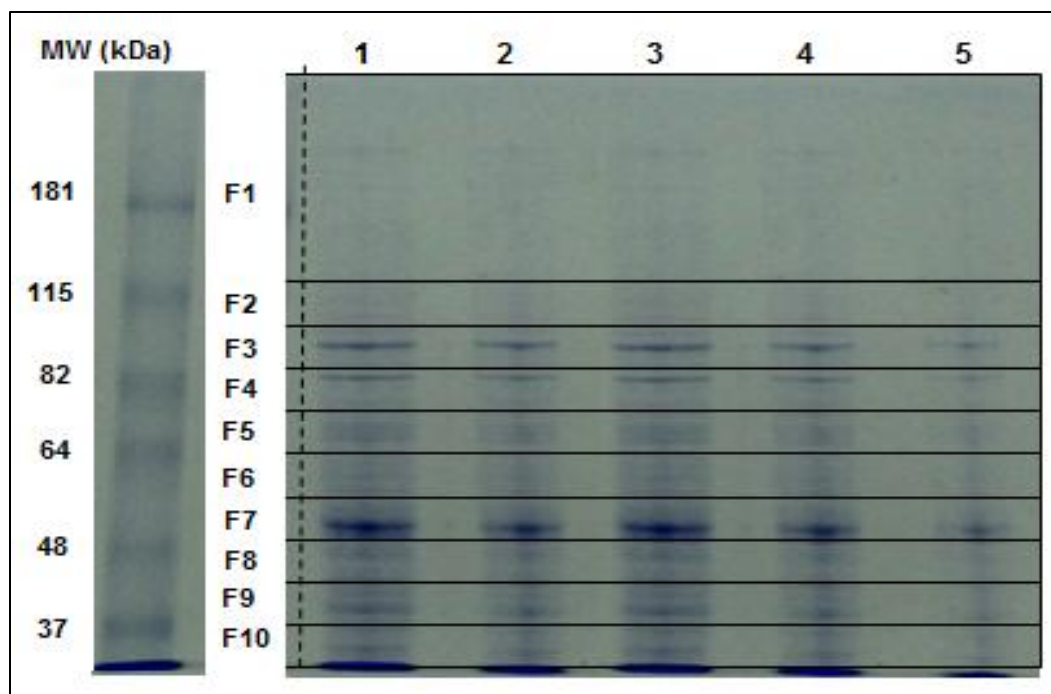
#### 3.4.6 Fracionamento e digestão de proteínas por gel gradiente SDS-PAGE

Este protocolo de fracionamento se baseou na separação das proteínas do extrato total de *T. cruzi* a partir de um gel SDS-PAGE com gradiente de 5 a 20% de poliacrilamida.

Foram preparados dois géis SDS-PAGE, um deles a 5% e outro a 20 % mantidos em gelo até a preparação do gel. Utilizada uma bomba peristáltica para transferir gradualmente e em paralelo cada gel até uma cuba de eletroforese vertical (Hoefer), a partir de uma base de vazamento onde cada gel é colocado, sob agitação, acoplado a uma mangueira (4 mm) chega até a cuba. Este aparato ficou sobre um agitador magnético, para evitar a polimerização antes da mistura de ambos os géis e evitar que o gradiente de concentração não seja formado adequadamente.

Após deixar o gel de separação depositar na cuba, foi adicionado butanol hidratado (1:1) sobre ele, após polimerização do gel, retirado o butanol e adicionado o gel de empilhamento (*stacking*), após polimerização foram aplicados 25 ug de extrato proteico total de *T. cruzi* Dm28c em 5 canaletas. Utilizado o marcador de massa molecular *Benchmark Protein Ladder* (Invitrogen). O gel foi corado com Coomassie Brilliant Blue G-250 e descorado com água (LAWRENCE; BESIR, 2009). A partir do gel foram geradas 10 frações cortando o gel horizontalmente de modo a manter em cada fração as proteínas pertencentes a uma determinada faixa de peso molecular (Figura 3.4).

Figura 3.4 - Modelo do fracionamento do gel gradiente SDS-PAGE



Fonte: o autor.

Dentro das frações (F1-F10) apenas as canaletas foram cortadas, as regiões entre elas foram descartadas. Por possuir menor número aparente de proteínas, F1 teve maior faixa de cobertura de peso molecular; MW, *Benchmark Protein Ladder* (Invitrogen).

As frações foram então submetidas à digestão de proteínas a partir de gel. Primeiramente, as frações de gel foram cortadas até próximo de 1x1 mm e separadas em tubos de microcentrífuga. Os géis foram descorados lavando duas vezes com solução de descoloração para Coomassie (bicarbonato de amônio 25 mM em etanol 50%) com quantidade suficiente para cobrir os pedaços de gel, agitando a 800 rpm no thermomixer por 10 minutos a 25 °C, e então, foram desidratados com etanol 100% nas mesmas condições. Em cada etapa retirado o sobrenadante e descartado.

Em seguida, os géis foram secados em Speed Vac (Thermo Scientific) por 7 minutos (*vacuum pressure* de 0,1). Reidratados os géis com tampão de redução (DTT 10 mM em ABC 50 mM), incubando a 800 rpm no thermomixer por 60 minutos a 56 °C. Retirado o sobrenadante e descartado. Após, foi adicionado tampão de alquilação (55 mM de iodoacetamida em 50 mM de ABC), incubando os géis no thermomixer a 800

rpm por 45 minutos a 25 °C, protegido da luz. Retirado o sobrenadante e descartado.

Na sequência, foi adicionado tampão de digestão (solução aquosa de ABC 50 mM) aos pedaços de gel, incubando em thermomixer a 800 rpm por 20 minutos a 25 °C e desidratados os géis com etanol 100%, nas mesmas condições. Repetidas estas etapas duas vezes. Os géis foram secados em Speed Vac (Thermo Scientific) por 7 minutos (*vaccum pressure* de 0,1). Após checar que os pedaços de gel estavam bem secos (“soltos” no tubo), foi adicionado solução de tripsina 12,5 ng/uL (diluída em ABC 50 mM no momento do uso), incubando os géis na geladeira (4 °C) por 20 minutos. Após, foi adicionado tampão de digestão aos géis e foram incubados na estufa a 37 °C por 16 horas.

Para interromper a digestão as amostras foram acidificadas com ácido trifluoracético (TFA), deixando-as em uma concentração final de 0,5%, agitadas em vortex por 10 segundos e coletado o líquido e transferido em tubos novos de microcentrífuga de 2,0 mL.

Em seguida, foi adicionada aos géis (não ao líquido dos tubos novos), a solução de extração (TFA 3% e ACN 30%), em quantidade suficiente para cobrir os géis, mas sem ultrapassar 400 uL e agitado no thermomixer a 800 rpm por 10 minutos a 25 °C. Coletado novamente o líquido e transferido aos tubos novos de 2,0 mL. Repetida essa etapa mais uma vez. Adicionado ACN 100% até cobrir os pedaços de gel, agitado em thermomixer a 800 rpm por 10 minutos a 25 °C e transferido o líquido para os tubos de 2,0 mL. Repetida essa etapa mais uma vez.

Por fim, as amostras foram secadas em Speed Vac (Thermo Scientific) a *vaccum pressure* de 0,1, até 10 a 20% do volume original, suficiente para remover a ACN. Os peptídeos foram purificados em *StageTips* C18 e analisados por LC-MS/MS.

### **3.5 Análise por nanoLC-MS/MS**

Após os fracionamentos os peptídeos foram separados por nanocromatografia líquida (nanoLC) de fase reversa seguida da ionização *eletrospray* (ESI) e análise por espectrometria de massas sequencial (MS/MS). O experimento foi realizado no sistema Easy-nLC 1000 (Thermo Fisher Scientific) conectado ao espectrômetro de massas

LTQ-Orbitrap XL ETD (Thermo Fisher Scientific) pela Plataforma de Espectrometria de Massas do Instituto Carlos Chagas.

Foram injetados 4 µg de amostra no sistema, provenientes de cada um dos fracionamentos realizados e de uma amostra peptídica também digerida por FASP, mas não fracionada, a fim de avaliar a eficiência do fracionamento. Após serem injetados, a separação peptídica foi feita em coluna analítica de 30 cm (diâmetro interno de 75 µm) empacotada pelo usuário com resina de fase reversa (partículas de C18 de 1,9 µm), com gradiente linear de 5 a 40% de acetonitrila (ACN) em 5% de DMSO, por 120 min. A coluna nanoLC foi aquecida a 60 °C para melhorar a resolução cromatográfica e reprodutibilidade. Os peptídeos foram ionizados por nanoeletrospray (voltagem de 2,3 kV) e injetados no espectrômetro de massas. Os espectros de massa foram obtidos por *Data-Dependent Acquisition* (DDA), com um *scan* inicial no Orbitrap com resolução R=15000 seguida por MS/MS dos 10 íons mais intensos no LTQ (*Linear Trap Quadrupole* ou *Iontrap*). Estes íons precursores foram fragmentados por dissociação induzida por colisão (*collision-induced dissociation*, CID) com energia de colisão normalizada de 35%, com tempo de ativação de 30 ms e ativação q = 0.25. Íons precursores com estados de carga <2 foram rejeitados. Paralelamente ao MS2, foi realizado um *full scan* no Orbitrap com uma resolução R=60.000 (amplitude de massa: m/z 300-2000) e os íons selecionados foram excluídos dinamicamente por 90 segundos. A opção *lock mass* foi utilizada em todos os *full scans* para obtenção de melhor acurácia de massa dos íons precursores.

### **3.6 Identificação dos peptídeos/proteínas em banco de dados**

A análise dos dados brutos (.raw), provenientes da análise LC-MS/MS, foi realizada através da plataforma MaxQuant (versão 1.5.2.8) (COX; MANN, 2008). Utilizou-se o banco de dados de sequências de proteínas do *T. cruzi* CL Brener do Uniprot ([www.uniprot.org](http://www.uniprot.org)), com um total de 19.242 sequências (04/01/2016). Proteínas contaminantes (queratinas humanas, BSA e tripsina porcina) foram adicionadas ao banco de dados, assim como suas sequências reversas. A busca pelos sítios metilados seguiu os seguintes parâmetros: tolerância no MS de 20 ppm



(Orbitrap), no MS/MS de 0.5 Da (Iontrap) permitindo até 2 sítios de clivagem perdidos (*missed cleavages*), carbamidometilação de cisteínas foi determinada como uma modificação fixa, assim como oxidação de metioninas. Os peptídeos foram buscados com no mínimo 7 aminoácidos. Além disso, um FDR (*False Discovery Rate*) de 1% foi aplicado em nível de proteína, peptídeo e metilsítio. Demais parâmetros empregados estão na Tabela 3.1.

Tabela 3.1 - Parâmetros utilizados na busca na plataforma MaxQuant

Parâmetro	
Fixed modifications	Carbamidomethyl (C)
Decoy mode	Revert
Special AAs	KR
Include contaminants	True
MS/MS tol. (FTMS)	20 ppm
Top MS/MS peaks per 100 Da. (FTMS)	12
MS/MS deisotoping (FTMS)	True
MS/MS tol. (ITMS)	0.5 Da
Top MS/MS peaks per 100 Da. (ITMS)	8
MS/MS deisotoping (ITMS)	False
MS/MS tol. (TOF)	40 ppm
Top MS/MS peaks per 100 Da. (TOF)	10
MS/MS deisotoping (TOF)	True
MS/MS tol. (Unknown)	0.5 Da
Top MS/MS peaks per 100 Da. (Unknown)	8
MS/MS deisotoping (Unknown)	False
PSM FDR	0.01
Protein FDR	0.01
Site FDR	0.01
Use Normalized Ratios For Occupancy	True
Min. peptide Length	7
Min. score for unmodified peptides	0
Min. score for modified peptides	40
Min. delta score for unmodified peptides	0
Min. delta score for modified peptides	6
Min. unique peptides	0
Min. razor peptides	1
Min. peptides	1
Use only unmodified peptides and	True
Modifications included in protein quantification	Acetyl (Protein N-term);Oxidation (M)

---

Peptides used for protein quantification	Razor
Discard unmodified counterpart peptides	True
Min. ratio count	2
Re-quantify	False
Use delta score	False
iBAQ	False
iBAQ log fit	False
Match between runs	True
Matching time window [min]	0.7
Alignment time window [min]	20
Find dependent peptides	False
Labeled amino acid filtering	True
Decoy mode	revert
Special AAs	KR
Include contaminants	True
RT shift	False
Advanced ratios	True
AIF correlation	0.47
First pass AIF correlation	0.8
AIF topx	20
AIF min mass	0
AIF SIL weight	4
AIF ISO weight	2
AIF iterative	True
AIF threshold FDR	0.01

---

Fonte: Maxquant.

### 3.7 Classificação funcional do metilproteoma de *T. cruzi*

As análises de classificação funcional e enriquecimento de GO (*Gene Ontology*) COG (*Clusters of Orthologous Groups*) e KEGG (*Kyoto Encyclopedia of Genes and Genomes*) e foram realizadas através da ferramenta *Database for Annotation, Visualization and Integrated Discovery* (DAVID) (HUANG; SHERMAN; LEMPICKI, 2009a, 2009b), versão 6.8 (<https://david.ncifcrf.gov>). Para as análises de enriquecimento, foi utilizado o FDR aplicado para a correção para múltiplos testes, conforme proposto por Benjamini e Hochberg (BENJAMINI; HOCHBERG, 1995) com *cutoff* menor ou igual a 0.05.

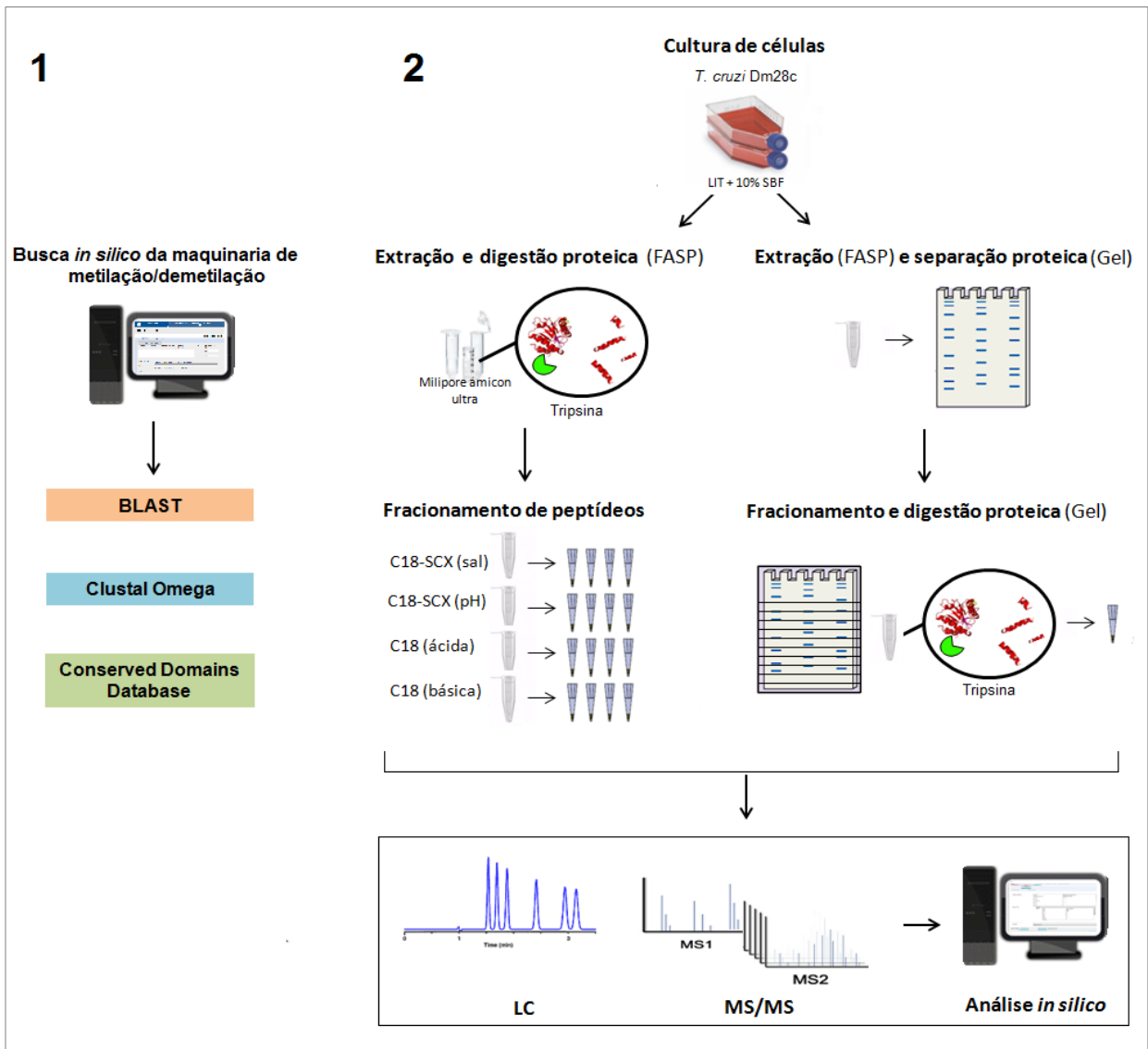
## 4 RESULTADOS

Para análise da maquinaria de metilação e caracterização do metilproteoma de *T. cruzi*, bem como de sua modulação ao longo da metaciclologênese, foi utilizado um conjunto de estratégias proteômicas e ferramentas bioinformáticas (Figura 4.1).

Metiltransferases e demetilases de *T. cruzi* e suas homólogas em *T. brucei* e humanos foram identificadas e caracterizadas *in silico* através de buscas na plataforma TriTrypDB; *Protein BLAST*, busca por domínios conservados através do algoritmo disponível na plataforma CDD e alinhamento de sequência primária para determinação de similaridade utilizando o algoritmo Clustal Omega (Figura 4.1-1).

Para identificação em larga escala de proteínas metiladas e sítios de metilação em formas epimastigotas de *T. cruzi*, células mantidas em cultura foram lisadas e as proteínas extraídas utilizando o protocolo FASP, e as amostras proteicas submetidas a diferentes protocolos de fracionamento. Os peptídeos, pós-digestão com tripsina, foram submetidos a cinco protocolos baseados em diferentes princípios de fracionamento: (i) C18-SCX (sal) - cromatografia por troca catiônica forte, em crescente concentração de  $\text{NH}_4\text{AcO}$ ; (ii) C18-SCX (pH) - cromatografia por troca catiônica forte em diferentes valores de pH; (iii) C18 ácida - cromatografia de fase reversa, pH ácido, com diferentes concentrações de solvente orgânico e (iv) C18 básica com frações concatenadas - cromatografia de fase reversa, pH básico. Com exceção da cromatografia de fase reversa em pH básico, os fracionamentos peptídicos foram realizados em microcolunas *StageTips*. Além dos fracionamentos a nível de peptídeos, o extrato proteico total foi submetido a fracionamento por (v) gradiente em gel SDS-PAGE para separação das proteínas íntegras por faixa de massa, seguido de digestão em gel com tripsina (Figura 4.1-2). Todas as amostras dos diferentes fracionamentos foram analisadas por LC-MS/MS e *softwares* proteômicos específicos.

Figura 4.1 - Desenho experimental empregado na caracterização do metilproteoma de *T. cruzi*.



Fonte: o autor.

Desenho experimental empregado na caracterização do metilproteoma de *T. cruzi*. (1) análise *in silico* das proteínas que compõem a maquinaria de metilação e demetilação em *T. cruzi*; (2) diferentes estratégias *bottom-up* na caracterização do metilproteoma da forma epimastigota de *T. cruzi*.

## 4.1 Análise, *in silico*, da maquinaria de metilação/demetilação de *T. cruzi*

### 4.1.1 Arginina metiltransferases de *T. cruzi* (TcPRMTs)

A busca por PRMTs de *T. cruzi* revelou a presença de 5 proteínas anotadas em ambos os haplótipos - *Esmeraldo-like* (E-L) e *non-Esmeraldo-like* (NE-L) - da cepa CL Brener, as quais possuem homólogos em *T. brucei* e *H. Sapiens* (Tabela 4.1).

Tabela 4.1 - TcPRMTs e seus homólogos em *T. brucei* e *H. sapiens*.

Organismo	PRMT	Uniprot ID	Gene IDs	Tipo de PRMT*	Número de aminoácidos	Mw (kDa)
<i>T. cruzi</i>	TcPRMT1	Q4E5S5	TcCLB.506529.50 (E-L)	I	441	51
		Q4DMV7	TcCLB.508593.110 (NE-L)		340	39
	TcPRMT3	Q4DTE9	TcCLB.510311.140 (E-L)	I	385	42
		Q4E5Y4	TcCLB.508153.1110 (NE-L)		385	42
	TcPRMT5	Q4D186	TcCLB.508781.129 (E-L)	II	377	42
		Q4D186	TcCLB.506853.30 (NE-L)		769	85
	TcPRMT6	Q4DA23	TcCLB.506947.80 (E-L)	I	374	42
		Q4CWZ1	TcCLB.507057.30 (NE-L)		374	42
	TcPRMT7	Q4DLS0	TcCLB.509153.100 (E-L)	III	419	46
		Q4DYZ1	TcCLB.510303.310 (NE-L)		419	46
<i>T. brucei</i>	TbPRMT1	Q4GYA9	Tb927.1.4690	I	345	39
	TbPRMT3	Q38BP3	Tb927.10.3560	I	389	42
	TbPRMT5	Q38CH6	Tb927.10.640	II	784	87
	TbPRMT6	Q57U70	Tb927.5.3960	I	368	41
	TbPRMT7	Q582G4	Tb927.7.5490	III	390	44
<i>H. sapiens</i>	HsPRMT1	Q99873	NP_938074	I	353	41
	HsPRMT3	O60678	NP_005779	I	531	60
	HsPRMT5	O14744	NP_006100	II	637	73
	HsPRMT6	Q96LA8	NP_060607	I	375	42
	HsPRMT7	Q9NVM4	NP_061896	II/III	692	78

Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*; E-L, *Esmeraldo-like*; NE-L, *non-Esmeraldo-like*. \*conforme Fisk; Read (2011) para *T. brucei* e *H. sapiens*; Fonte: o autor.

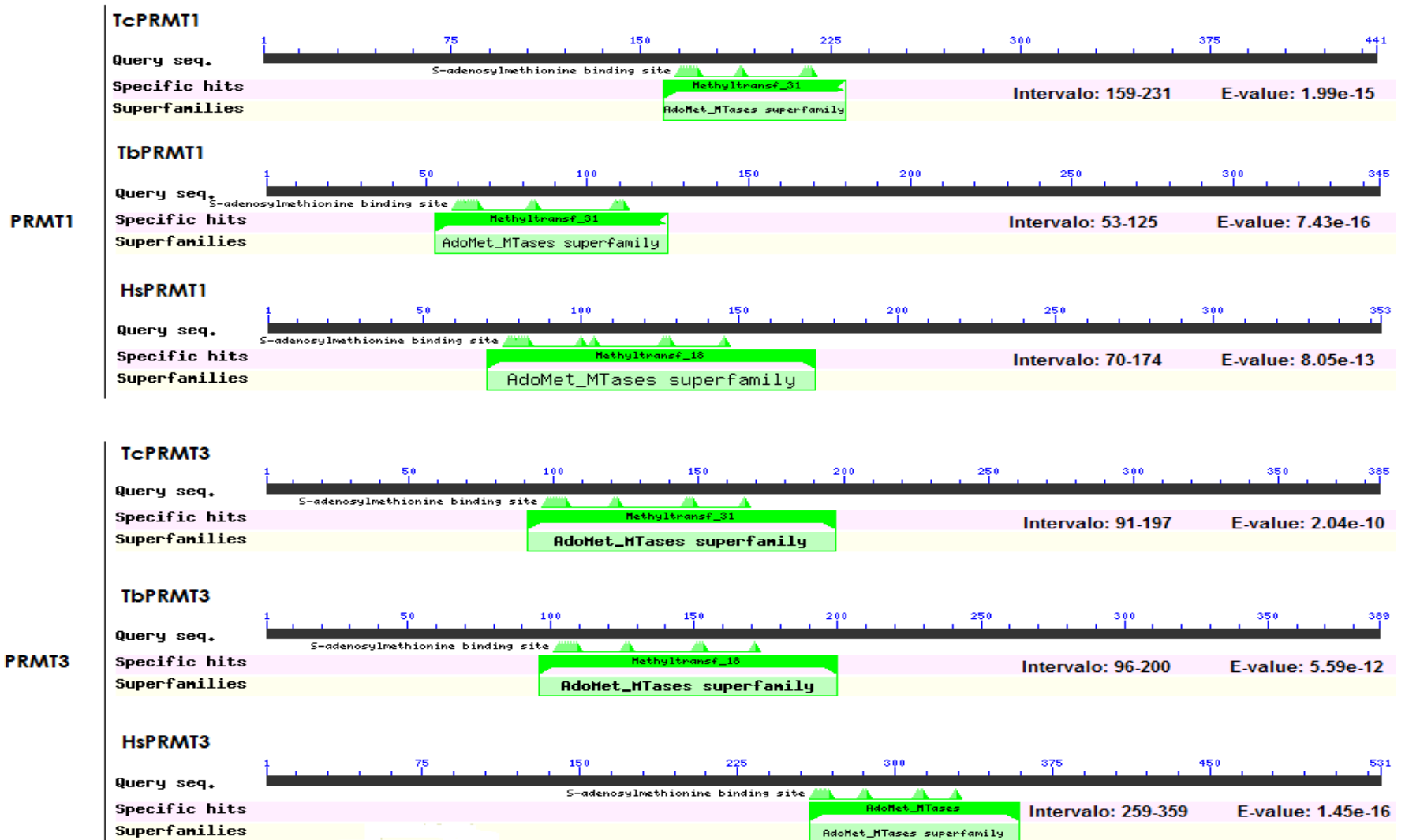
As PRMTs anotadas para *T. cruzi* variam entre 39 e 85 kDa e apresentam pequenas variações de tamanho em relação às suas proteínas homólogas em *T. brucei* e humanos. As TcPRMTs 1 e 5 apresentam dois genes anotados com tamanhos diferentes, sendo um com tamanho similar e outro maior ou menor em relação às homólogas em *T. brucei* e humanos, respectivamente. As PRMTs 3 e 7 têm tamanho similar em *T. cruzi* e *T. brucei*, sendo ambas menores que suas homólogas em humanos. Já a PRMT6 apresenta tamanho similar nos três organismos.

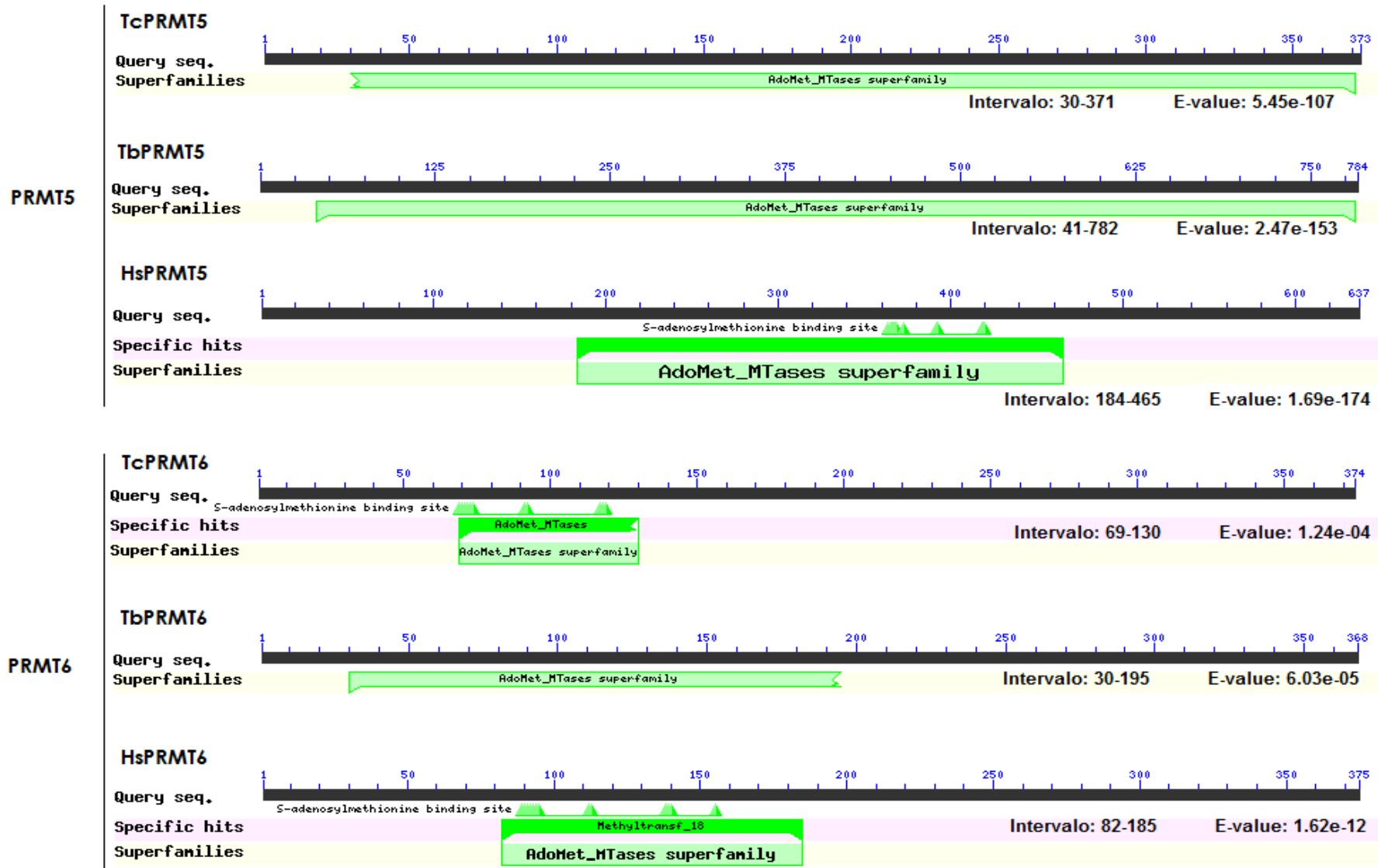
As sequências dos haplótipos E-L e NE-L anotados para cada TcPRMT específica mostraram-se bastante similares em todos os casos (Anexo 1). Para as PRMTs 3, 6 e 7 a identidade de sequência está entre 97,4 e 98,5%. Já para as PRMTs 1 e 5, devido à diferença de tamanho entre os genes, a identidade global ficou em torno de 75% e 46%, respectivamente. Entretanto, nas regiões presentes em ambos os haplótipos, tanto a PRMT1 quanto a PRMT5 apresentaram alta identidade, como observado para as outras. Assim, para as demais análises, foram utilizadas as sequências do haplótipo E-L, considerada cepa referência.

Foi então realizada a busca pela presença de domínios conservados nestas sequências de *T. cruzi*, bem como em suas homólogas em *T. brucei* e humanos. Todas as proteínas foram associadas (*e-value* abaixo de 0.01) com a família das metiltransferases dependentes de AdoMet as quais utilizam essa molécula como substrato doador de grupos metil durante o processo de metilação de seus alvos. Quase todas apresentam regiões de sequência com alta associação (specific hits) a domínios destas metiltransferases (Figura 4.2).

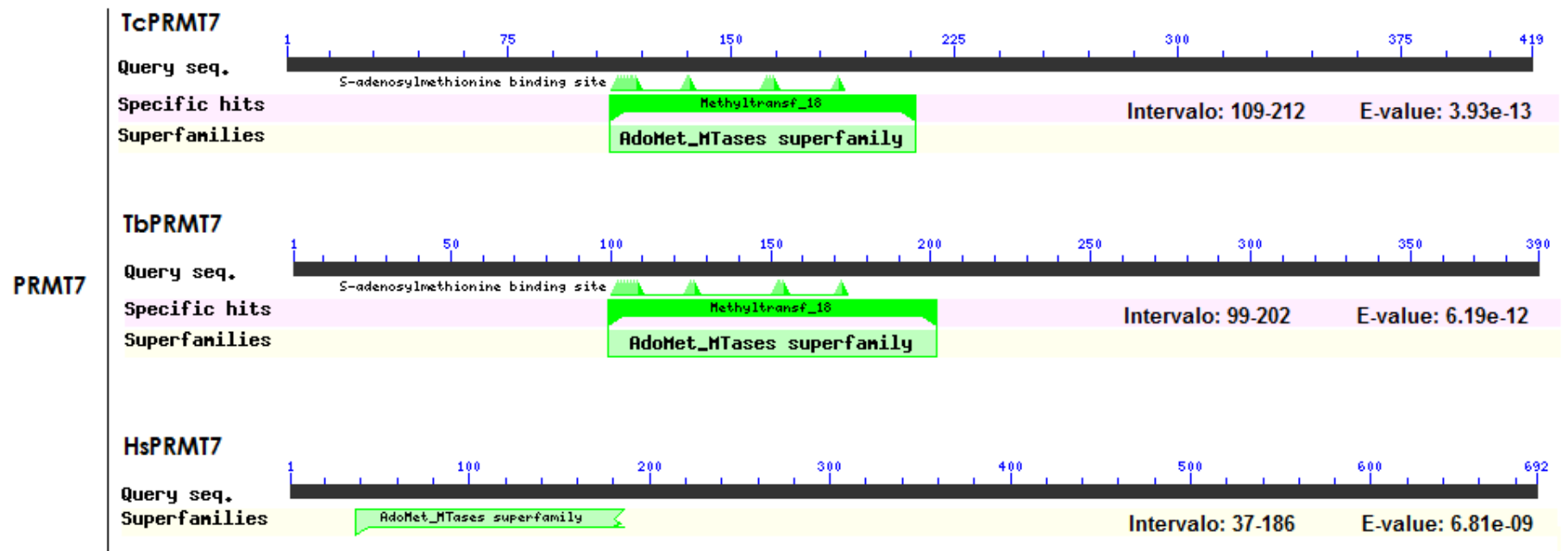
Em seguida, as sequências das TcPRMTs foram alinhadas com suas homólogas em *T. brucei* e humanos, e apresentaram identidade de 34%, 14,6%, 10%, 19% e 9,5% para as PRMTs 1, 3, 5, 6 e 7, respectivamente. Apesar da identidade geral de sequência não ser tão alta entre os três organismos, o que em parte pode ser reflexo de diferenças de tamanho, a região do domínio metiltransferase apresenta-se bastante conservada, especialmente nos resíduos que promovem a ligação da enzima às moléculas de AdoMet. Como esperado, a maior identidade está presente nas regiões de sequência contendo os domínios: 68,3%; 33,3%; 20%; 27% e 28,1% entre as PRMT1, PRMT3, PRMT5, PRMT6 e PRMT7, respectivamente (Figura 4.3).

Figura 4.2 – Domínios conservados das PMRTs de *T. cruzi*, *T. brucei* e *H. sapiens*.



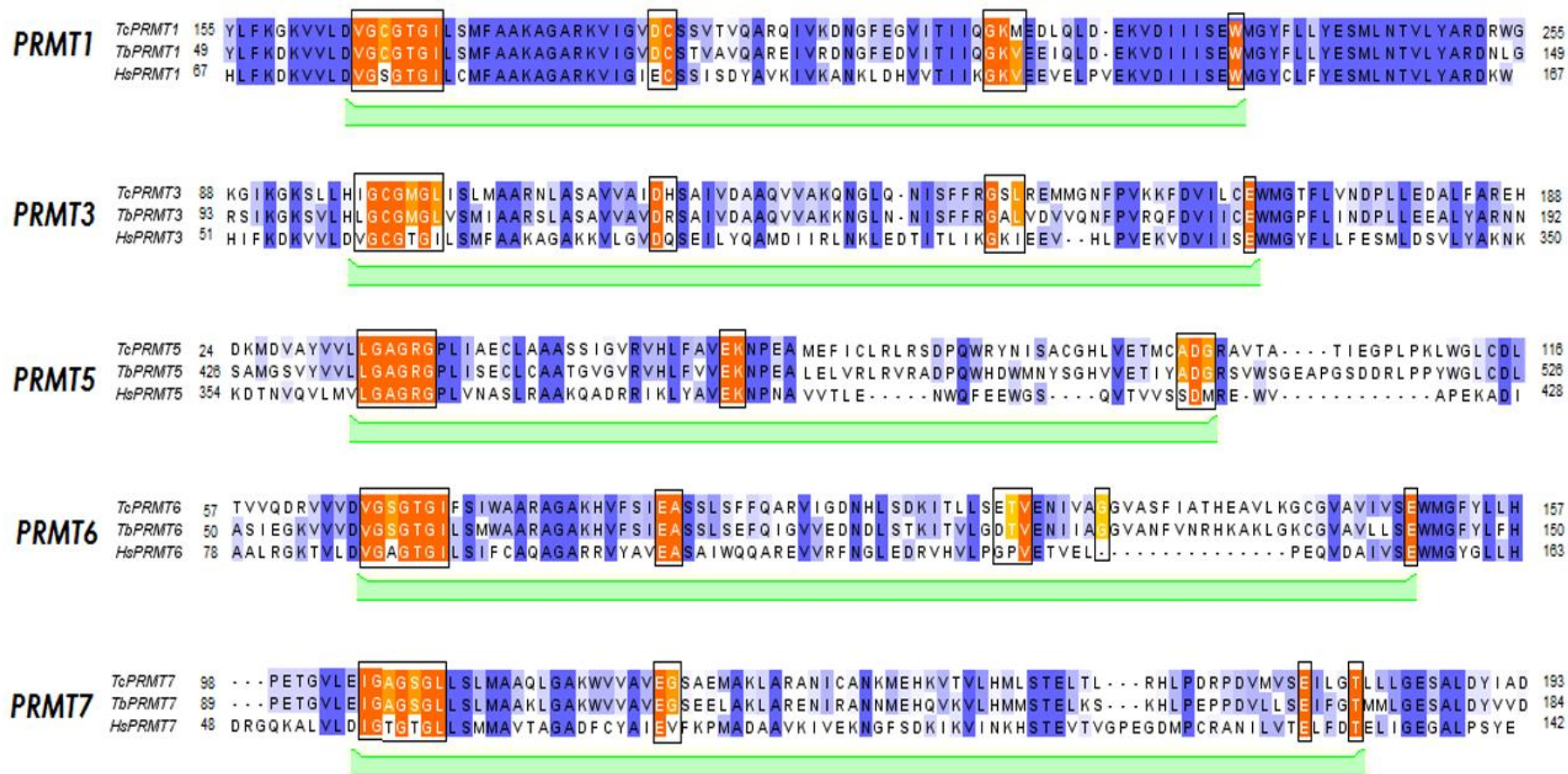






A figura inclui os membros do haplótipo E-L de *T. cruzi* e seus respectivos homólogos em *T. brucei* e humanos, listados na Tabela 4.1. Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em verde, os domínios conservados das metiltransferases dependentes de AdoMet.

Figura 4.3 - Alinhamento da região do domínio catalítico das PRMTs de *T. cruzi*, *T. brucei* e *H. sapiens*.



O alinhamento inclui os membros do haplótipo E-L de *T. cruzi* e seus respectivos homólogos em *T. brucei* e humanos, listados na Tabela 4.1. Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em tons de azul e laranja, diferentes níveis de conservação entre os aminoácidos, onde tons escuros indicam aminoácidos idênticos e tons mais claros indicam aminoácidos similares. A barra verde indica a região do sítio catalítico das metiltransferases dependentes de AdoMet. Retângulos pretos indicam as regiões conservadas que se ligam a AdoMet.

#### 4.1.2 Lisina metiltransferases de *T. cruzi* (TcPKMTs) - Família DOT1

Para verificar a presença de PKMTs em *T. cruzi*, foram buscadas na plataforma TritypDB as sequências já reportadas como lisina N-metiltransferases. Foram encontradas 3 PKMTs da família DOT1, todas anotadas para o haplótipo E-L, as quais possuem homólogos em *T. brucei* e *Homo sapiens*, conforme determinado através de BLAST. Pertencem à PKMT4 (ZHANG, WEN, SHI, 2012), conhecida classe de PKMTs sem o domínio SET (Tabela 4.2).

Tabela 4.2 - TcPKMTs e seus homólogos em *T. brucei* e *H. sapiens*.

Organismo	PKMT	Uniprot ID	Gene ID	Tipo de PKMT	Número de aminoácidos	Mw (kDa)
<i>T. cruzi</i>	TcDOT1B <sup>+</sup>	Q4D9P2	TcCLB.511417.70 (E-L)	PKMT4	275	31
	TcDOT1 <sup>++</sup>	Q4DXY7	TcCLB.510431.70 (E-L)	PKMT4	301	34
		Q4CKM5	TcCLB.460125.10 (E-L)			218
<i>T. brucei</i>	TbDOT1A	Q581Z0	Tb927.8.1920	PKMT4	295	33
	TbDOT1B	Q4GZF2	Tb927.1.570	PKMT4	275	31
<i>Homo sapiens</i>	HsDOT1	Q8TEK3	NP_115871	PKMT4*	1537	165

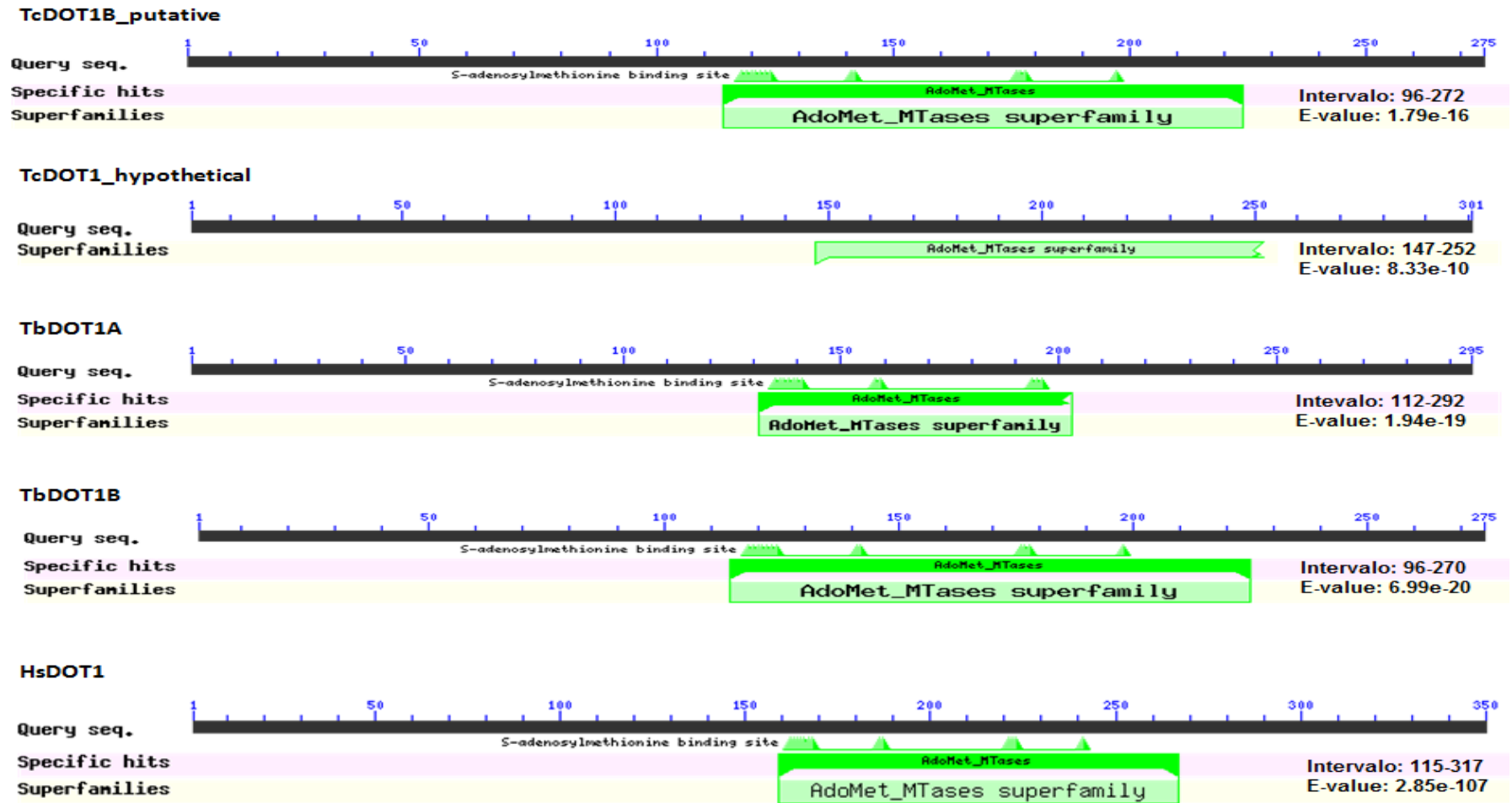
<sup>+</sup>Putativa; <sup>++</sup>Hipotética; E-L, *esmeraldo-like*. \* conforme Zhang, Wen, Shi (2012)

As PKMTs de *T. cruzi* variam entre 31 e 34 kDa (o gene de 24 kDa está anotado como possível fragmento no TriTrypDB), apresentando tamanho similar à suas homólogas em *T. brucei*, mas bastante menor que as homólogas humanas. Todas as TcPKMTs foram consideradas para as análises posteriores.

A partir de buscas por domínios conservados, foram identificados motivos típicos de metiltransferases dependente de AdoMet em todas as sequências de TcPKMTs, assim como em suas homólogas de *T. brucei* e humanos (Figura 4.3). Em seguida, as sequências das PKMTs homólogas nos três organismos foram alinhadas para verificar as regiões conservadas. As PKMTs DOT1B (putativa) de *T. cruzi* e *T. brucei* apresentaram identidade de 57%, indicando que estas proteínas podem ser consideradas homólogas nesses organismos. Já em relação à proteína de humanos, TcDOT1B apresenta baixa identidade (3%) entre estas sequências decorrente da grande diferença de tamanho entre elas, nas regiões dos domínios ligantes a AdoMet são encontradas regiões bem conservadas (Figura 4.4). Isso ocorre mesmo entre as sequências TcDOT1B (putativa) e TcDOT1 (hipotética)

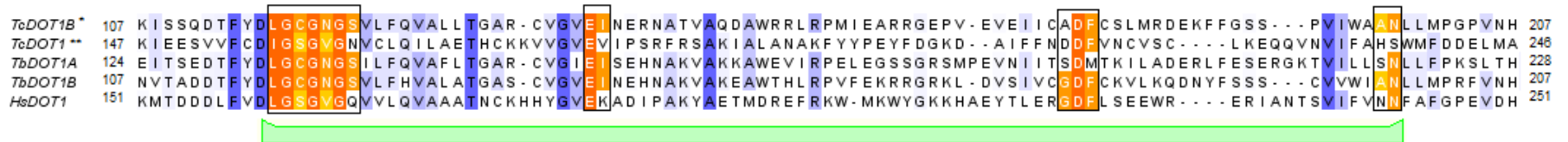
quando alinhadas com HsDOT1 pois a identidade que era de apenas 3% passa para 16% e 14%, respectivamente, na região onde estão presentes os domínios identificados.

Figura 4.4 – Domínios conservados das PKMTs da família DOT1 de *T. cruzi*, *T. brucei* e *H. sapiens*.



A figura inclui todos os membros da família DOT1 de *T. cruzi*, *T. brucei* e humanos listados na Tabela 4.2. Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em verde, os domínios conservados das metiltransferases dependentes de AdoMet.

Figura 4.5 - Alinhamento da região do domínio catalítico das PKMTs da família DOT1 em *T. cruzi*, *T. brucei* e *H. Sapiens*.



O alinhamento inclui todos os membros da família DOT1 de *T. cruzi*, *T. brucei* e humanos listados na Tabela 4.2. Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em tons de azul e laranja, diferentes níveis de conservação entre os aminoácidos, onde tons escuros indicam aminoácidos idênticos e tons mais claros indicam aminoácidos similares. A barra verde indica a região do sítio catalítico das metiltransferases dependentes de AdoMet. Retângulos pretos indicam as regiões conservadas que se ligam a AdoMet.

#### 4.1.3 Demetilases contendo domínios JmjC de *T. cruzi*

Para concluir a busca pela maquinaria de metilação, foi verificada a presença de demetilases contendo o domínio JmjC em *T. cruzi*. Foram buscadas, na plataforma TritypDB, as sequências já reportadas com o domínio JmjC e encontradas 7 possíveis demetilases, todas anotadas como proteínas hipotéticas conservadas, sendo 2 delas para o haplótipo E-L e 5 para o haplótipo NE-L. Dentre as demetilases putativas de *T. cruzi*, 5 possuem homólogos em *T. brucei* e *Homo sapiens*, 1 possui homólogo apenas em *T. brucei* e 1 apenas com *H. sapiens*, conforme determinado através de BLAST (Tabela 4.3).

Tabela 4.3 - Demetilases de *T. cruzi* e seus homólogos em *T. brucei* e *H. sapiens*

Organismo	Demetilase	Uniprot ID	Gene ID	Tipo de demetilase	Número de aminoácidos	Mw (kDa)
<i>T. cruzi</i>	TcJmjD4	Q4D046	TcCLB.506369.30 (E-L)	JHDM	428	48
		Q4CXU5	TcCLB.510225.50 (NE-L)	JHDM	451	51
	TcJmjD6	Q4E398	TcCLB.508355.310 (NE-L)	JHDM	555	63
	TcJmjD7	Q4DVZ6	TcCLB.506289.60 (E-L)	JHDM	553	63
		Q4DMI3	TcCLB.508881.30 (NE-L)	JHDM	472	53
	TcNO66	Q4D641	TcCLB.508141.30 (NE-L)	JHDM	476	54
	n.d.	Q4DVQ0	TcCLB.510659.130 (NE-L)	JHDM	1155	129
<i>T. brucei</i>	TbJmjD4	Q386X8	Tb927.11.1760	JHDM	450	50
	TbJmjD6	Q386V9	Tb927.11.2000	JHDM	550	62
	TbJmjD7	Q57VR2	Tb927.7.660	JHDM	466	53
	n.d.	Q38DD6	Tb927.9.12780	JHDM	1145	126
<i>H. sapiens</i>	HsJmjD4	Q9H9V9	NP_075383.2	JHDM*	463	52
	HsJmjD6	Q6NYC1	NP_055982.2	JHDM*	403	46
	HsJmjD7	P0C870	NP_001108104.1	JHDM*	316	36
	HsNO66	Q9H6W3	NP_078920.2	JHDM*	641	71

JHDMs, JmjC-domain-containing histone demethylases; n.d., não determinado; E-L, *Esmeraldo-like*; NE-L, *non-Esmeraldo-like*, Mw, *molecular weight*; \*(KLOSE; KALLIN; ZHANG, 2006)

As demetilases de *T. cruzi* variam entre 48 e 129 kDa, dentre as que possuem sequência similar com *T. brucei* e humanos, a massa molecular fica entre 48 e 63 kDa, ou seja, tamanhos bastante similares entre as que apresentaram homologia.

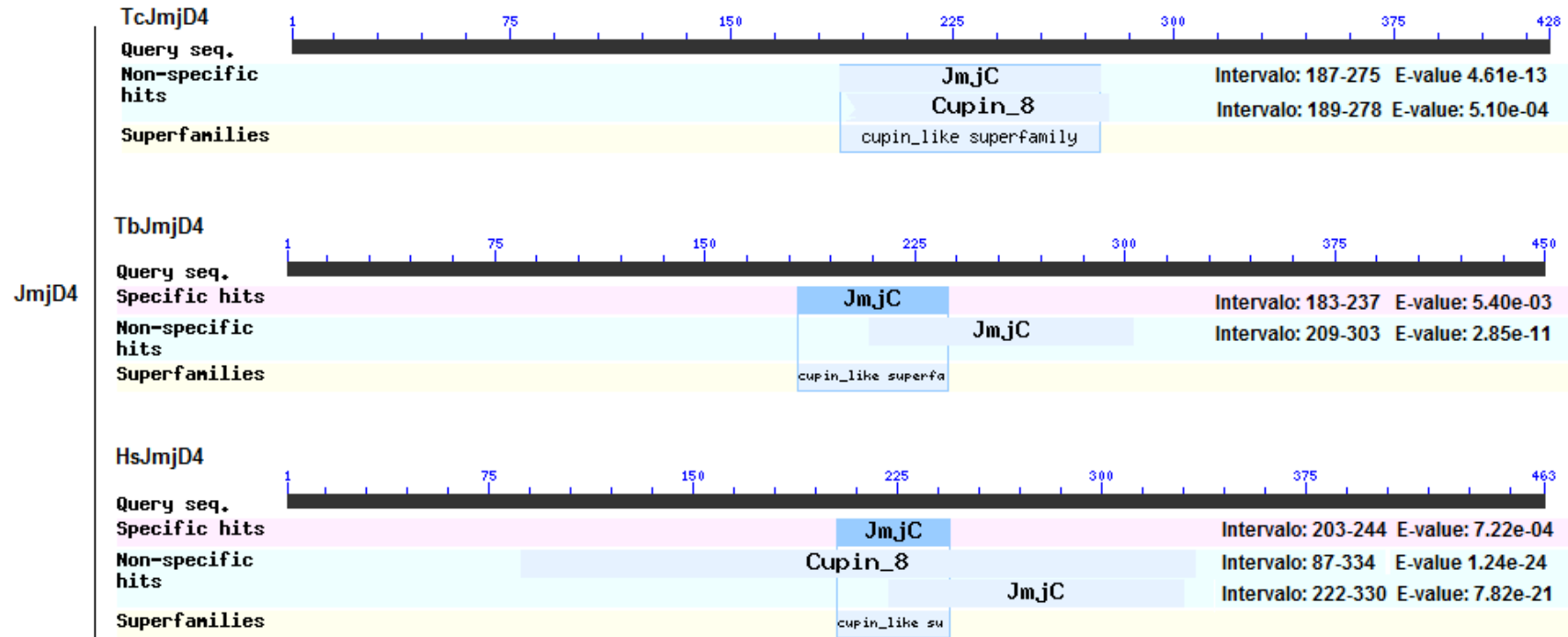
As sequências de *H. sapiens* e *T. brucei* pertencem às demetilases de histona que contêm domínio JmjC (*JmjC-domain-containing histone demethylases*, JHDM) (KLOSE; KALLIN; ZHANG, 2006). As proteínas de *T. cruzi* apresentam-se como proteínas hipotéticas conservadas, no entanto, por homologia e presença de domínios (Figura 4.6) foram inferidas como possíveis JHDMS.

A não ser pela presença do domínio JmjC nas proteínas TcCLB.510659.130 de *T. cruzi* e Tb927.9.12780 de *T. brucei* não foi possível inferir a que demetilase esse gene é homólogo com humano.

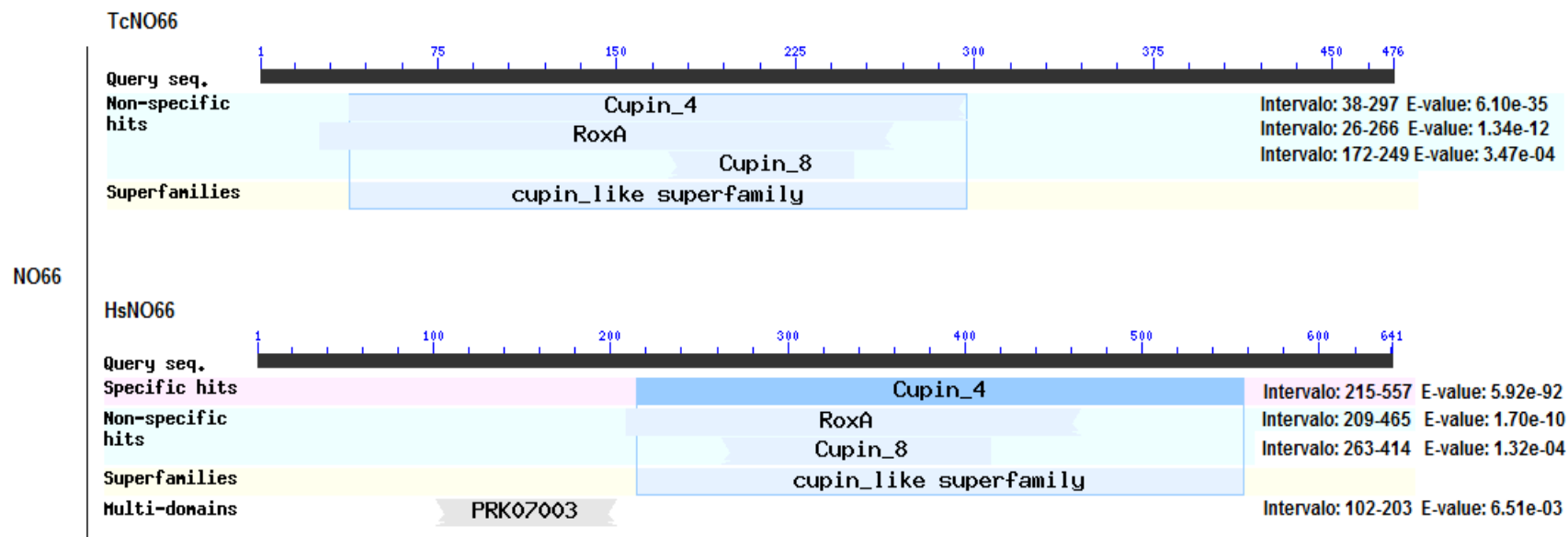
O alinhamento das regiões dos domínios entre as sequências da Tabela 4.3 está na Figura 4.7. As sequências dos haplótipos E-L e NE-L que estão anotadas para as demetilases de *T. cruzi* mostraram-se bastante similares (Anexo 1) e para este alinhamento foram utilizadas as sequências E-L quando ambos os haplótipos estavam presentes (Tabela 4.3). A identidade apresentada na região do sítio catalítico dentro do domínio das demetilases entre *T. cruzi*, *T. brucei* e humanos foi de 31,8% para JmJD4, 29,7% para JmjD6, 26,8% para JmjD7 e 49,3% para NO66 (esta entre *T. cruzi* e humanos). Dentro dessa região estão presentes os resíduos que caracterizam o domínio JmjC, os quais são o sítio de ligação a  $\alpha$ -cetogluturato ( $\alpha$ KG-*binding*) e ao Ferro II (Fe(II)-*binding*) (CHANG et al., 2007; KLOSE; KALLIN; ZHANG, 2006). Para a sequência de NO66 não foi encontrada similaridade significativa através de BLAST em *T. brucei*.



Figura 4.6 - Domínios conservados das demetilases com domínio JmjC em *T. cruzi*, *T. brucei* e *H. sapiens*.



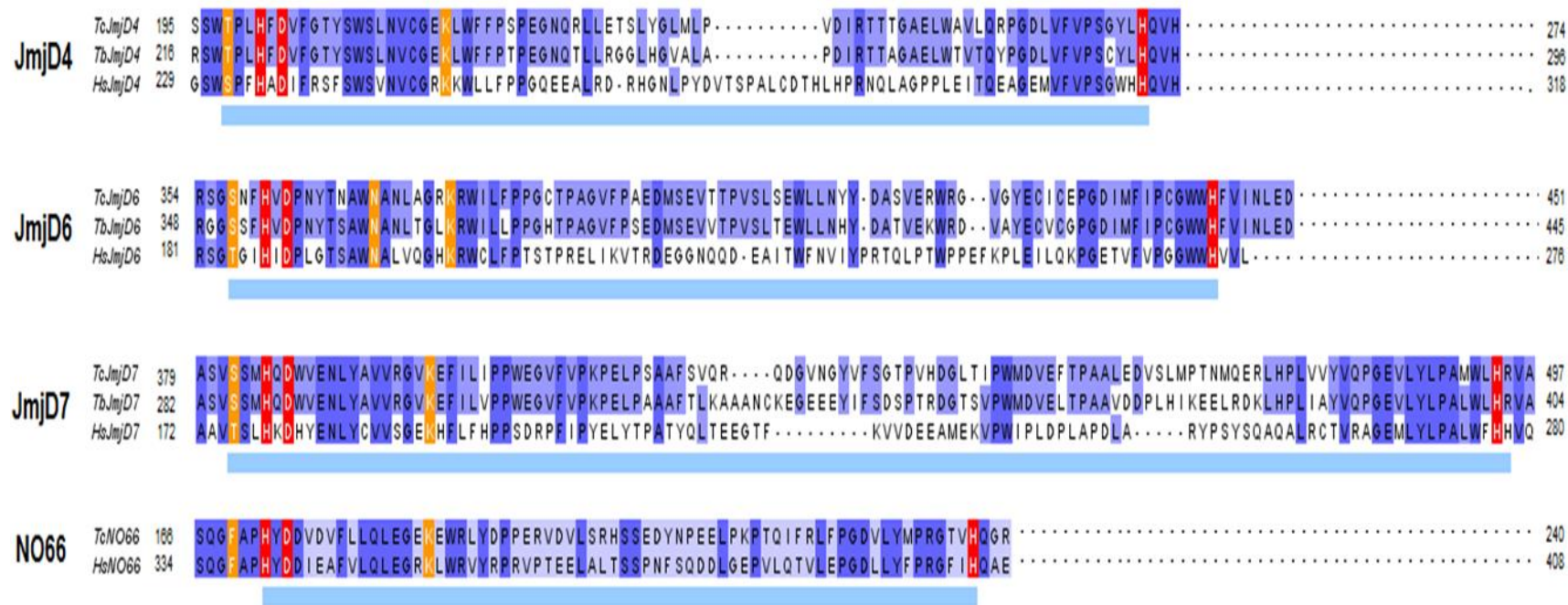




Fonte: o autor.

A figura inclui todos os membros das demetilases de *T. cruzi*, *T. brucei* e humanos listados na Tabela 4.3. Tc, *T. cruzi* CL Brener E-L, exceto em TcJmjD6 (NE-L); Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em azul, os domínios conservados das demetilases com domínio JmjC, o qual pertence à família Cupin.

Figura 4.7 - Alinhamento da região do domínio catalítico das demetilases de *T. cruzi*, *T. brucei* e *H. sapiens*.



Fonte: o autor.

O alinhamento inclui os membros do haplótipo E-L de *T. cruzi* (exceto JmjD6 que o haplótipo é NE-L) e seus respectivos homólogos em *T. brucei* e humanos, listados na Tabela 4.3. Tc, *T. cruzi* CL Brener E-L; Tb, *T. brucei* TREU927; Hs, *Homo sapiens*. Em tons de azul, diferentes níveis de conservação entre os aminoácidos, onde tons escuros indicam aminoácidos idênticos e tons mais claros indicam aminoácidos similares. A barra azul indica a região do sítio catalítico das demetilases onde estão os resíduos  $\alpha$ KG-binding (laranja) e Fe(II)-binding (vermelho) que caracterizam o domínio JmjC. *T. brucei* não apresenta para NO66 uma sequência similar às sequências de *T. cruzi* e *H. sapiens*.

## 4.2 Caracterização do metilproteoma de formas epimastigotas de *T. cruzi*

### 4.2.1 Análise preliminar, *in silico*, do metilproteoma de *T. cruzi*

Para confirmar a ocorrência dos diferentes tipos de metilação, determinar o potencial papel da metilação na regulação de processos biológicos, foi inicialmente realizada uma análise preliminar buscando sítios de metilação de *T. cruzi*, a partir da reanálise de arquivos gerados em trabalho proteômico anterior (DE GODOY et al., 2012).

Foram identificados 4760 grupos proteicos distintos, incluindo 158 proteínas metiladas. Dentre as proteínas metiladas foram encontradas 65 (41,2%) com sítios de Me, 87 (55%) com sítios de Me<sub>2</sub> e 6 (3,8%) com sítios de Me<sub>3</sub>. Algumas proteínas apresentaram mais de um tipo de metilação. Por exemplo, 6 proteínas trimetiladas também estavam dimetiladas; 3 proteínas dimetiladas também apresentaram sítios de monometilação e uma proteína monometilada encontra-se trimetilada.

Dentre as 158 proteínas metiladas identificadas, 81 (51%) delas são anotadas em pelo menos alguma das três categorias de ontologia gênica (ASHBURNER et al., 2000). Os dados indicam que as proteínas metiladas representam diferentes componentes celulares, tais como citoplasma (10%); membrana (5%), incluindo tanto as proteínas integrais de membrana quanto da membrana de organelas celulares; retículo endoplasmático (1%); núcleo (1%); mitocôndria (1%) e Golgi (1%). Em termos de processo biológico a categoria mais representada corresponde as proteínas envolvidas no metabolismo celular (11%), principalmente atuando na síntese e degradação de componentes celulares, como de aminoácidos, seguida da categoria com proteínas envolvidas no folding proteico (6%), muitas delas chaperoninas. Outra classe representada está envolvida com o transporte (4%), incluindo componentes da cadeia pesada de clatrina e subunidade da importina, proteína envolvida no transporte de outras proteínas para dentro do núcleo celular.

#### 4.2.2 Eficiência de diferentes protocolos de fracionamento na identificação de proteínas metiladas

Após a comprovação *in silico* da presença da maquinaria de metilação e a identificação de proteínas metiladas em dados já publicados do grupo, foram avaliados diferentes protocolos para a caracterização em larga escala do metilproteoma de *T. cruzi*.

Visando melhorar a extração de proteínas, foi utilizado o protocolo de FASP, o qual permite a lise celular na presença de SDS, melhorando a solubilização e recuperação de proteínas que normalmente são perdidas ou sub-representadas em outros métodos de extração mais fracos, tais como proteínas integrais de membrana, associadas a ácidos nucléicos, etc. A fim de reduzir a complexidade e assim aumentar a identificação de peptídeos e sítios de metilação de *T. cruzi* na etapa de LC-MS/MS, as amostras protéicas foram submetidas a cinco tipos diferentes de fracionamento (Figura 3.1-2). Os números resultantes de cada protocolo testado estão sumarizados na Tabela 4.4.

Tabela 4.4 - Proteínas, peptídeos e sítios identificados nos diferentes fracionamentos

Fracionamento	Proteínas		Peptídeos		Sítios					
	Total	Metiladas	Total	Metilados	R	K	Me	Me2	Me3	Total
C18 (ácida)	1900	119	7658	115	96	58	58	80	16	154
C18-SCX (sal)	3736	305	16753	300	185	238	249	127	47	423
C18 (básica)	2897	326	12273	329	121	321	173	176	93	442
C18-SCX (pH)	2812	310	9111	319	253	246	227	195	77	499
Gel gradiente SDS-PAGE	4133	878	20577	965	657	679	700	507	129	1336

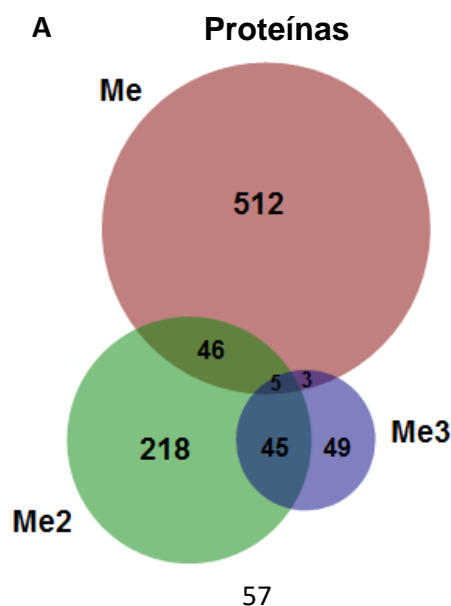
Fonte: o autor.

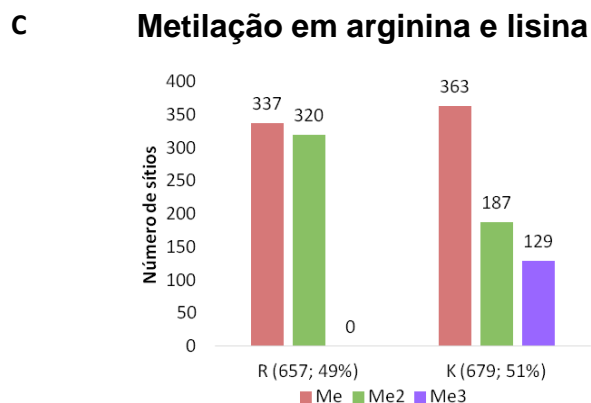
Dentre os diferentes fracionamentos testados, o que demonstrou maior eficiência foi a separação do extrato total de *T. cruzi* em gel gradiente SDS-PAGE, com posterior divisão das canaletas em faixas de massa molecular, digestão de proteínas em gel e análise por LC-MS/MS. Assim, esse conjunto de dados foi selecionado para a realização das análises posteriores referentes ao metilproteoma de formas epimastigotas de *T. cruzi*.

#### 4.2.3 O metilproteoma de formas epimastigotas de *T. cruzi*

A análise por LC-MS/MS das 10 frações do gel SDS-PAGE do extrato total de *T. cruzi*, resultou na aquisição de um total de 116405 espectros MS/MS. Estes espectros foram confrontados com o banco de dados e 51323 espectros (cerca de 44%) foram identificados. Após a exclusão dos contaminantes e sequências reversas identificadas, foi obtido um total de 20577 peptídeos e 4133 proteínas. Do total de proteínas identificadas na forma epimastigota de *T. cruzi*, 878 proteínas estavam metiladas (Tabela S1) e com um total de 1336 sítios de metilação (Me/Me2/Me3) (Tabela S2). A distribuição dos diferentes tipos de metilação está apresentada na Figura 4.8. A maioria das proteínas metiladas apresentaram sítios de monometilação (58,3%), seguidas daquelas com sítios de dimetilação (24,8%) e trimetilação (5,6%). Algumas metilproteínas apresentaram, ainda, mais de um tipo de metilação. (Figura 4.8 A). Dentre os sítios de metilação, 700 (52%) representam monometilação (363 Kme, 337 Rme), 507 (38%) dimetilação (187 Kme, 320 Rme) e 129 (10%) trimetilação (Figura 4.8 B). Os sítios de metilação mostraram-se quase igualmente distribuídos entre resíduos de arginina (49%) e lisina (51%) (Figura 4.8 C). Do total de sítios identificados, 1263 (94,5%) possuem alta probabilidade de localização (*Localization Probability* >0.75).

Figura 4.8 - Distribuição Me/Me2/Me3 dos eventos de metilação em epimastigotas de *T. cruzi*





Fonte: o autor.

A) Diagrama de Venn mostrando a sobreposição dos tipos de metilação (Me, Me2 e Me3) nas proteínas metiladas identificadas; B) Distribuição dos sítios de metilação identificados quanto ao tipo de metilação (Me, Me2 e Me3); C) Distribuição dos sítios de metilação (Me, Me2 e Me3) nos aminoácidos arginina (R) e lisina (K)

#### 4.2.3.1 Panorama geral da localização e função das proteínas reguladas por metilação em epimastigotas de *T. cruzi*

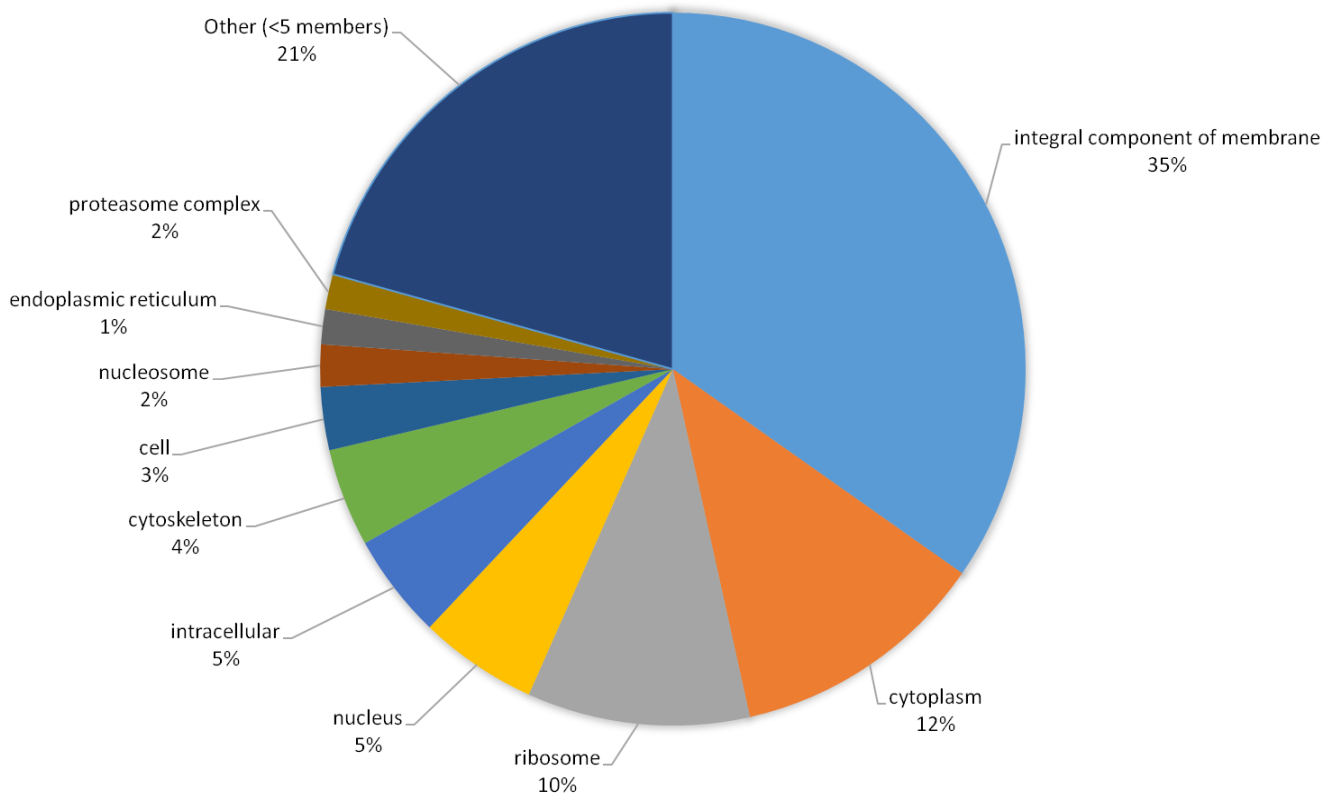
Para determinar as vias potencialmente reguladas pelo metilproteoma (metilação em arginina e lisina) na forma epimastigota de *T. cruzi*, foram realizadas as análises de classificação funcional e enriquecimento dos termos de GO das 878 proteínas metiladas identificadas. A primeira análise consistiu na classificação das proteínas metiladas de acordo com as anotações GO, COG e KEGG para *T. cruzi* CL Brener. Para apresentação dos resultados, classes contendo 5 proteínas ou mais, ou seja, as mais representadas dentro do nosso conjunto de 878 proteínas, foram ilustradas individualmente, sendo as demais (menos de 5 proteínas) agrupadas na classe *others*. As listas completas de classificação e enriquecimento estão disponíveis nas Tabela S3 e Tabela S4, respectivamente.

A classificação de GO a partir dos termos de componente celular (*Cellular Component*) revelou que 285 (32,4%) das proteínas metiladas identificadas possuem anotação para esta categoria, distribuídas em 59 diferentes classes (Tabela S3), 10 delas com 5 ou mais proteínas. Como pode ser visto na Figura 4.9, deste subconjunto 109 proteínas estão na classe componente integral de membrana, 37 são proteínas citoplasmáticas, 32 são proteínas associadas aos ribossomos, 17 ao núcleo e 14 ao



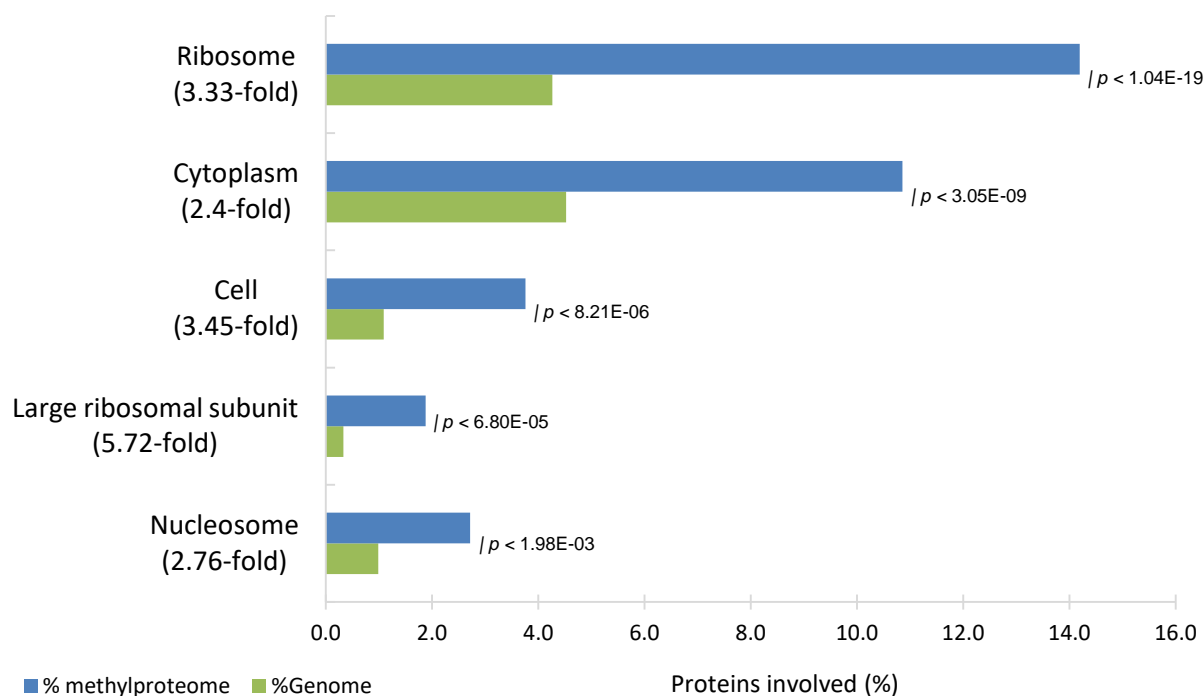
citoesqueleto, entre outras classes. Já a análise de enriquecimento demonstrou que as proteínas metiladas identificadas estão diretamente envolvidas com ribossomos, representado pelas classes *Ribosome* e *Large ribosomal subunit* (Figura 4.10). Além dessas duas classes, está bem representada no conjunto das metiladas as proteínas relacionadas ao nucleossomo, a unidade estrutural dos cromossomos. Destaca-se, portanto, para este termo de GO analisado o envolvimento das proteínas metiladas com componentes celulares relacionados à transcrição e, principalmente, síntese de proteínas.

Figura 4.9 - Proteínas metiladas de *T. cruzi* classificadas de acordo com Componente Celular.



Fonte: o autor.

Figura 4.10 - Componentes celulares enriquecidos dentre as proteínas metiladas de *T. cruzi*.



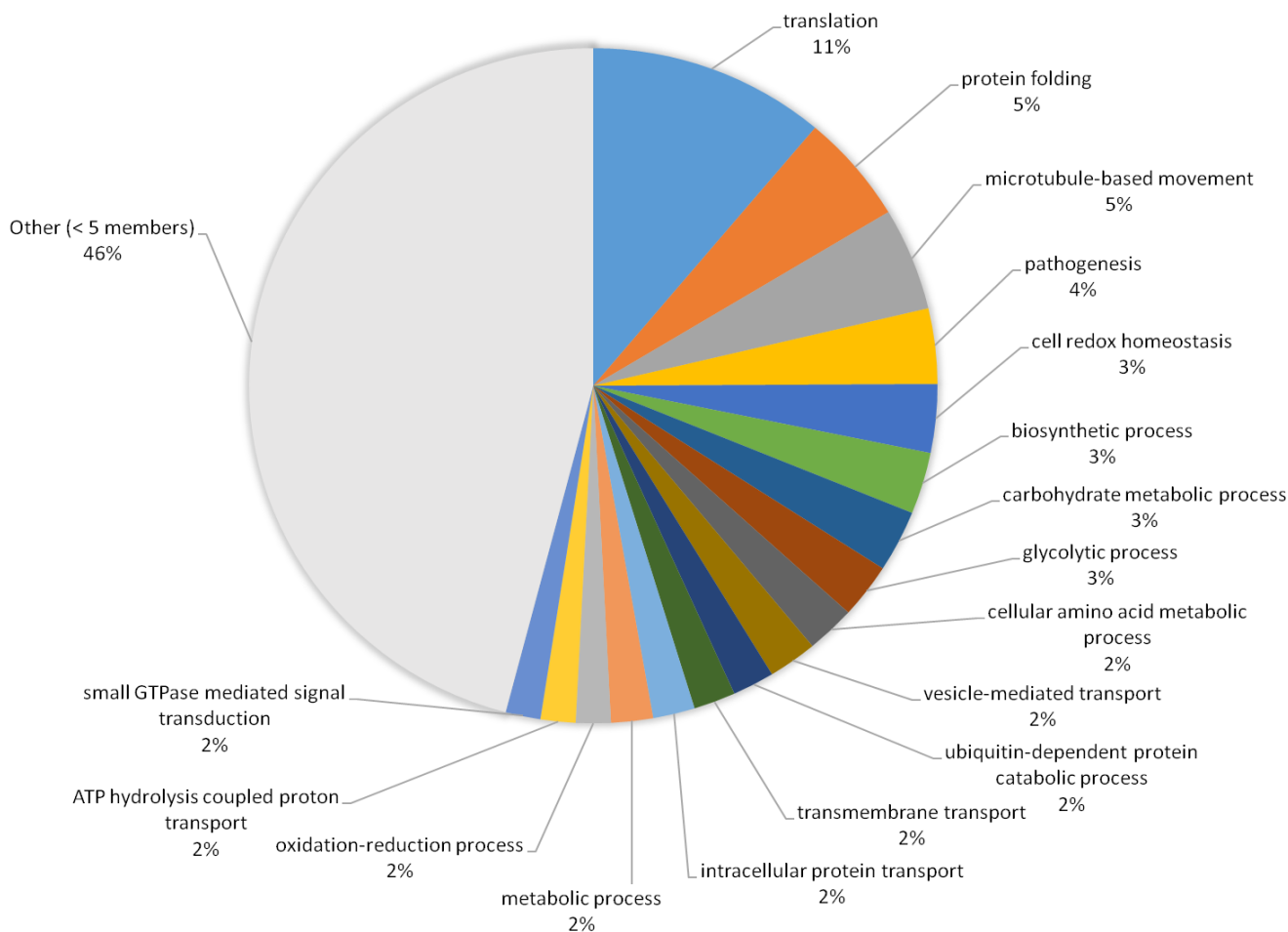
Fonte: o autor

Para a categoria de Processo Biológico (*Biological Process*), 295 (33,6%) das proteínas metiladas identificadas possuem termos anotados e estão distribuídas em 143 diferentes classes (Tabela S3), sendo 17 com cinco ou mais proteínas (Figura 4.11). A classe mais representada possui 34 proteínas envolvidas com o processo de tradução, também aparece significativamente enriquecida (Figura 4.12). Em outras classes estão 16 proteínas envolvidas no *fold*ing proteico, 15 ao movimento por microtúbulos e 11 ao processo patogênico do parasito. A homeostasia redox da célula também está representada por duas classes, *cell redox homeostasis* com 10 proteínas e *oxidation-reduction process* com mais 5 proteínas. Outras classes bem representadas estão relacionadas a processos metabólicos celulares, como o metabolismo de carboidratos representado por 9 (*carbohydrate metabolic process*) e 8 (*glycolytic process*) proteínas e o transporte intracelular, onde aparece o mediado por vesículas com 7 proteínas (*vesicle-mediated transport*) e o transporte de

proteínas com 6 proteínas.

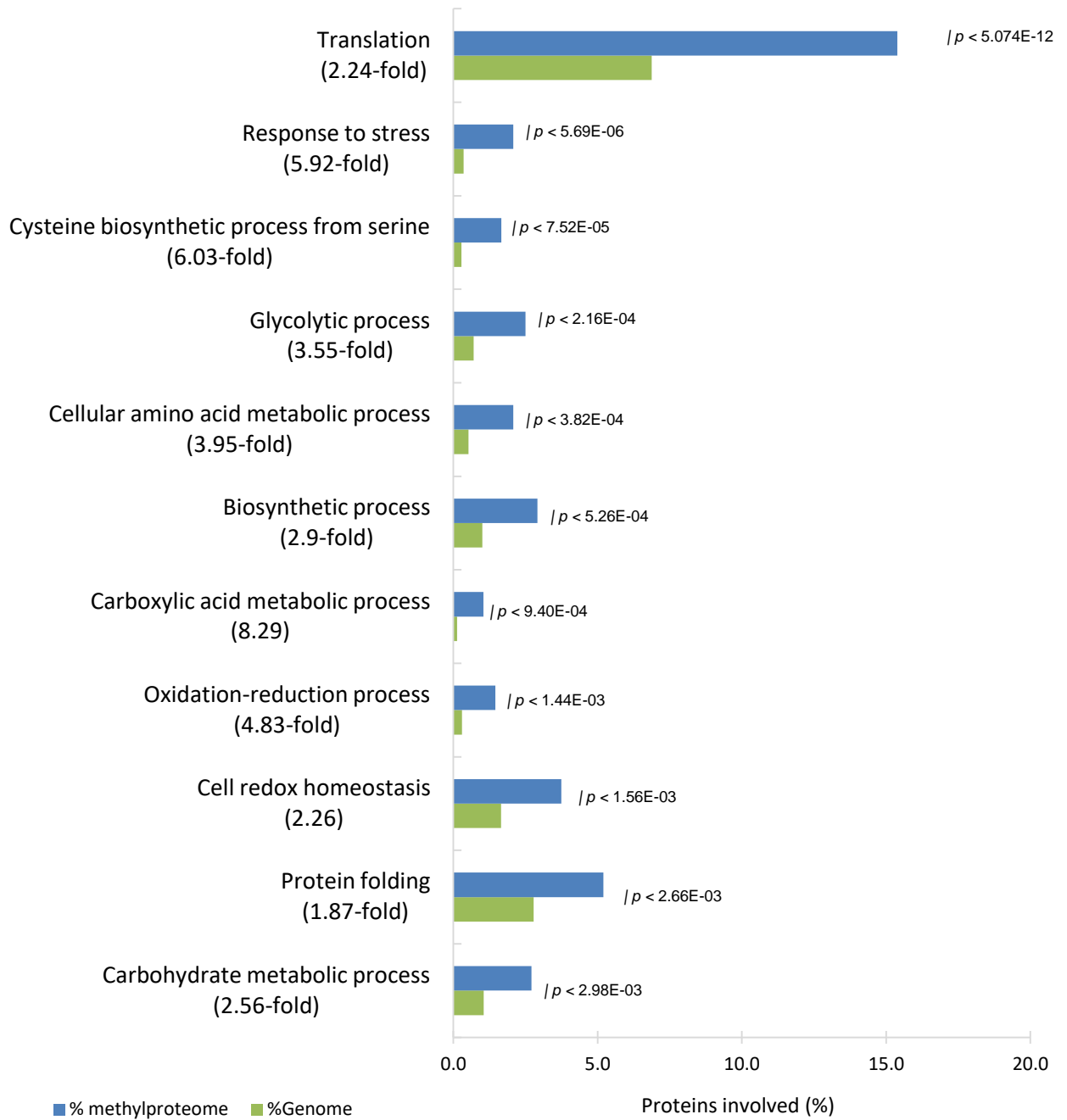
O metilproteoma de *T. cruzi* apresentou 11 processos biológicos significativamente enriquecidos (Figura 4.12). Dentre estas categorias, o processo de tradução, que apresenta-se como o mais representado em número de proteínas dentro da classificação, também está entre as classes enriquecidas. Processo metabólico dos ácidos carboxílicos, aparece como o processo biológico mais enriquecido.

Figura 4.11 – Proteínas metiladas de *T. cruzi* classificadas de acordo com Processo Biológico.



Fonte: o autor.

Figura 4.12 – Processos biológicos enriquecidos dentre as proteínas metiladas de *T.cruzi*.

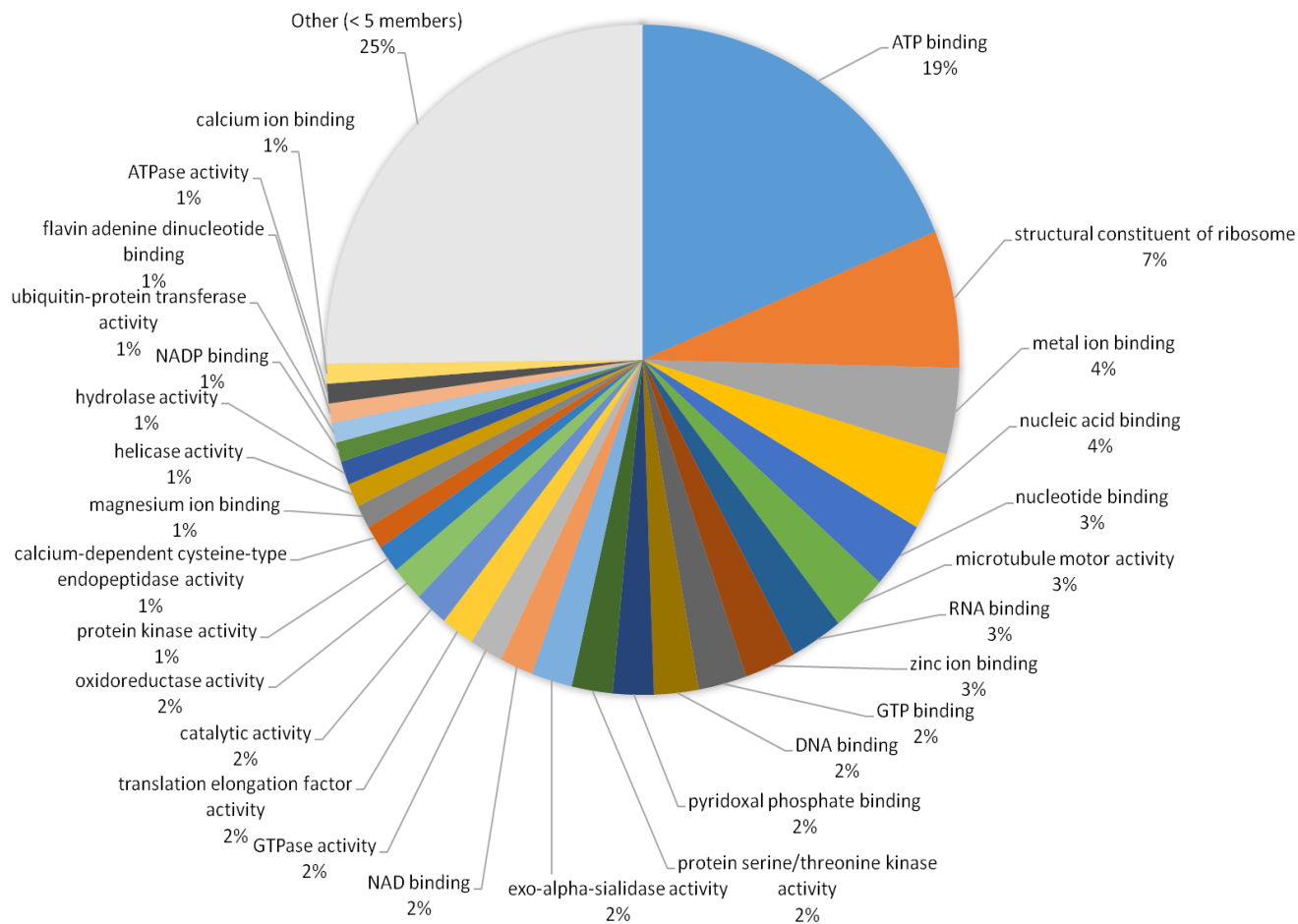


Fonte: o autor.

Outra categoria de GO empregada na classificação foi a Função Molecular (Figura 4.13). Nesta categoria, 415 (47,2%) proteínas estavam anotadas e amplamente distribuídas em 210 classes (Tabela S3), incluindo 28 classes com 5 ou mais proteínas. Destas, a maior parte delas, 99 proteínas, estão na classe *ATP binding*, seguido de *structural constituent of ribosome*, com 35 proteínas. Proteínas envolvidas com a ligação de íons, representadas nas categorias *metal ion binding*, *zinc ion binding*, *magnesium ion binding*, *calcium ion binding*, somam 47 proteínas. Outra importante classe representada é a das proteínas envolvidas com ligação à ácidos nucleicos, cujas classes *nucleic acid binding*, *RNA binding* e *DNA binding* representam 46 proteínas. Interessantemente, 18 proteínas metiladas identificadas estão anotadas com a função molecular de quinase nas classes *protein serine/threonine kinase activity* e *protein kinase activity* com 11 e 7 proteínas, respectivamente.

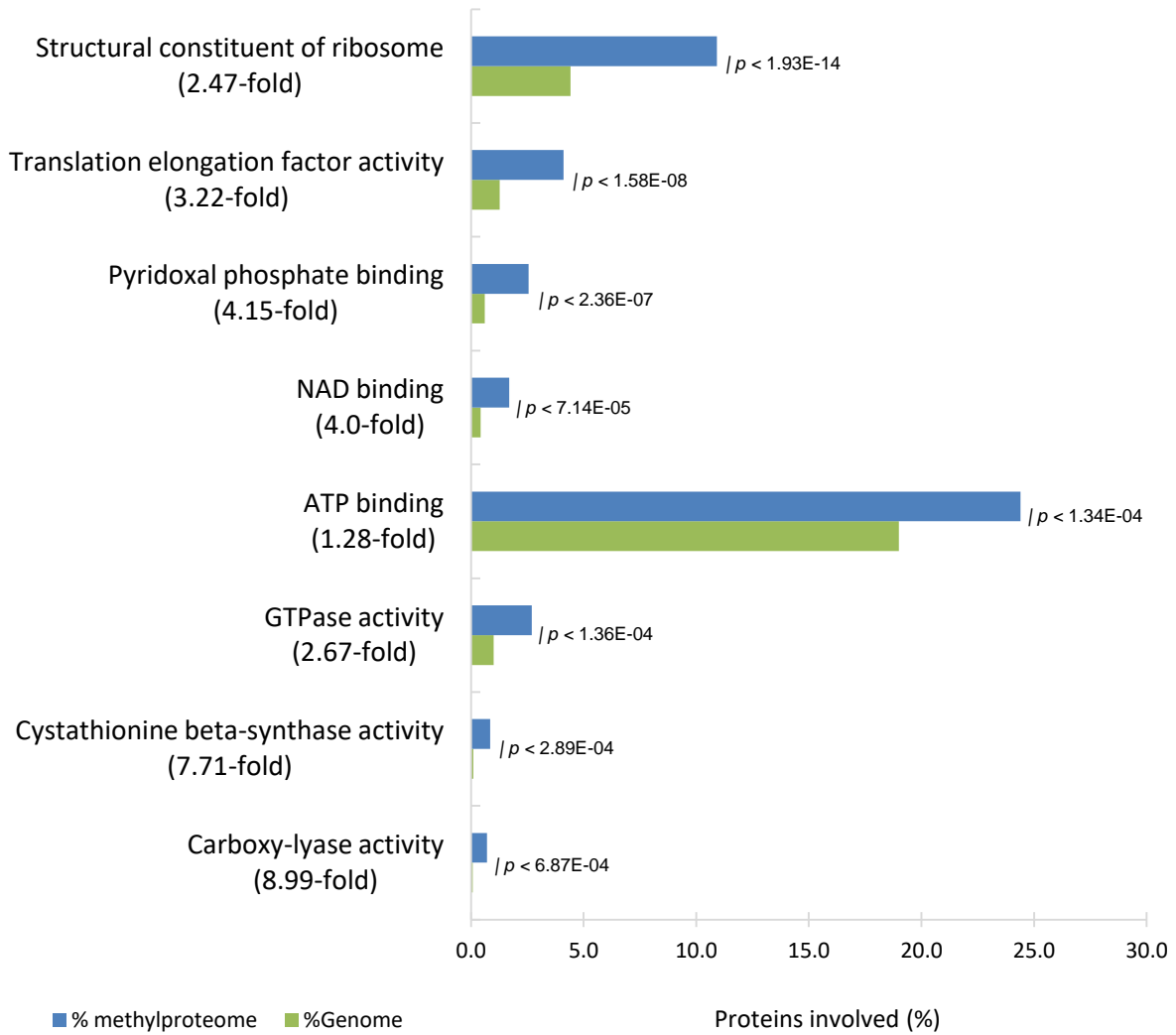
Apesar do grande número de proteínas metiladas relacionadas à função *ATP binding*, essa classe não aparece como uma das mais enriquecidas dentro do genoma de *T. cruzi* (Figura 4.14), outras funções como *translation elongation factor activity* é representado por 9 proteínas, dentre elas *elongation factor 1-alpha* (EF-1-alpha) (Tc00.1047053511367.360), *elongation factor 1-gamma* (EF-1-gamma) (Tc00.1047053510163.20), *elongation factor 2* (Tc00.1047053508169.20), *transcription elongation factor-like protein* (Tc00.1047053507715.30).

Figura 4.13 – Proteínas metiladas de *T. cruzi* classificadas de acordo com Função Molecular.



Fonte: o autor

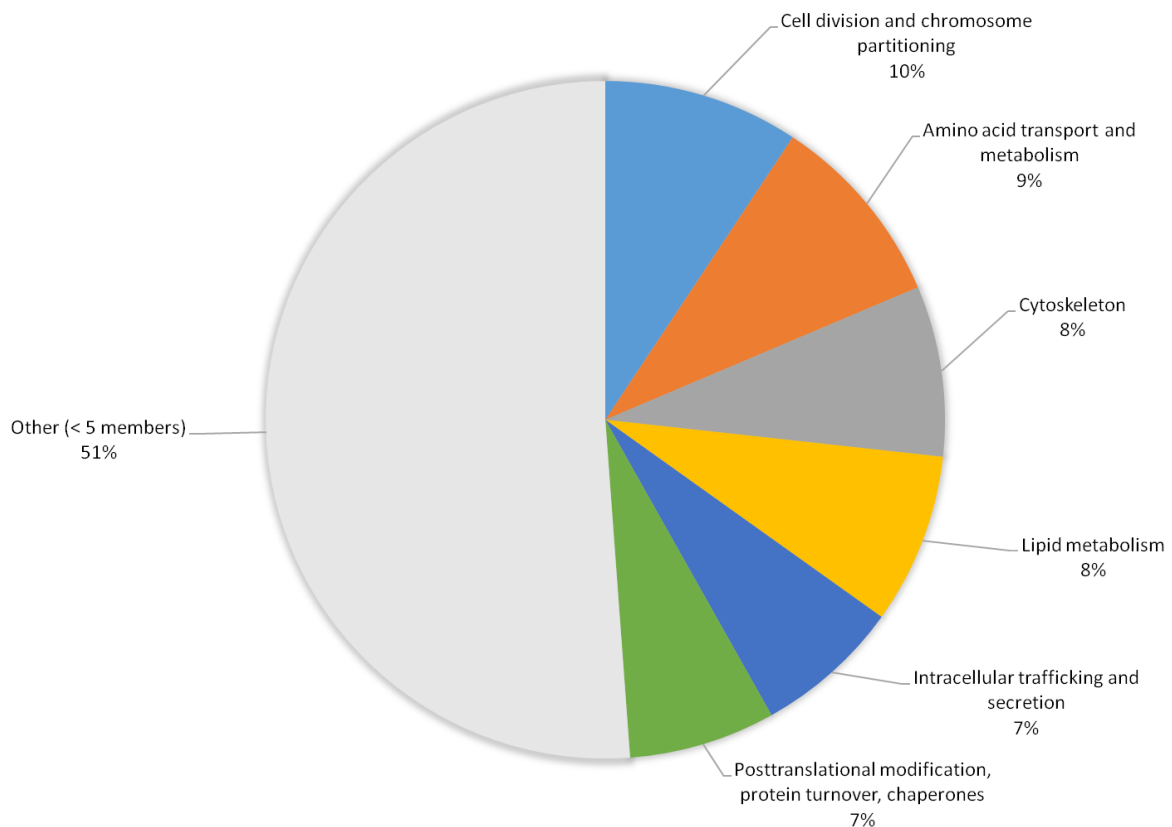
Figura 4.14 – Funções moleculares enriquecidas dentre as proteínas metiladas de *T.cruzi*.



Fonte: o autor.

A classificação por *Clusters of Orthologous Group* (COG) foi recuperada para 86 (9,8%) proteínas, distribuídas em 20 classes (Tabela S3). Os 6 grupos, onde os mais representados ( $\geq 5$  proteínas) foram *cell division and chromosome partitioning* e *amino acid transport and metabolism*, com 8 proteínas cada, seguidos de *Cytoskeleton* e *Lipid metabolism*, com 7 proteínas cada e *Intracellular trafficking and secretion* e *posttranslational modification, protein turnover, chaperones* com 6 proteínas cada (Figura 4.15). Para a categoria de COG, não foram encontradas classes enriquecidas dentre as proteínas metiladas.

Figura 4.15 – Proteínas metiladas de *T. cruzi* classificadas de acordo com COG.

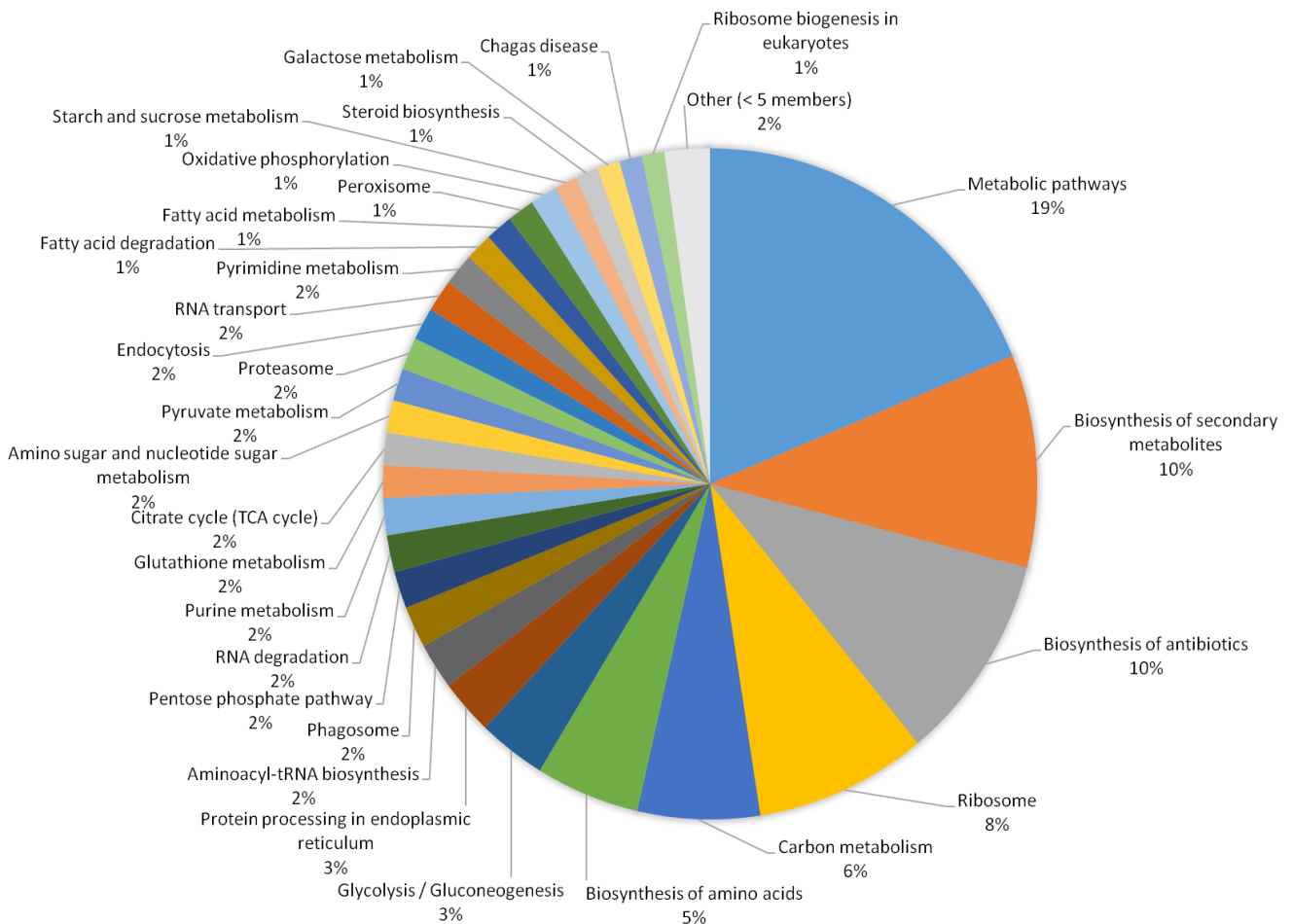


Fonte: o autor.



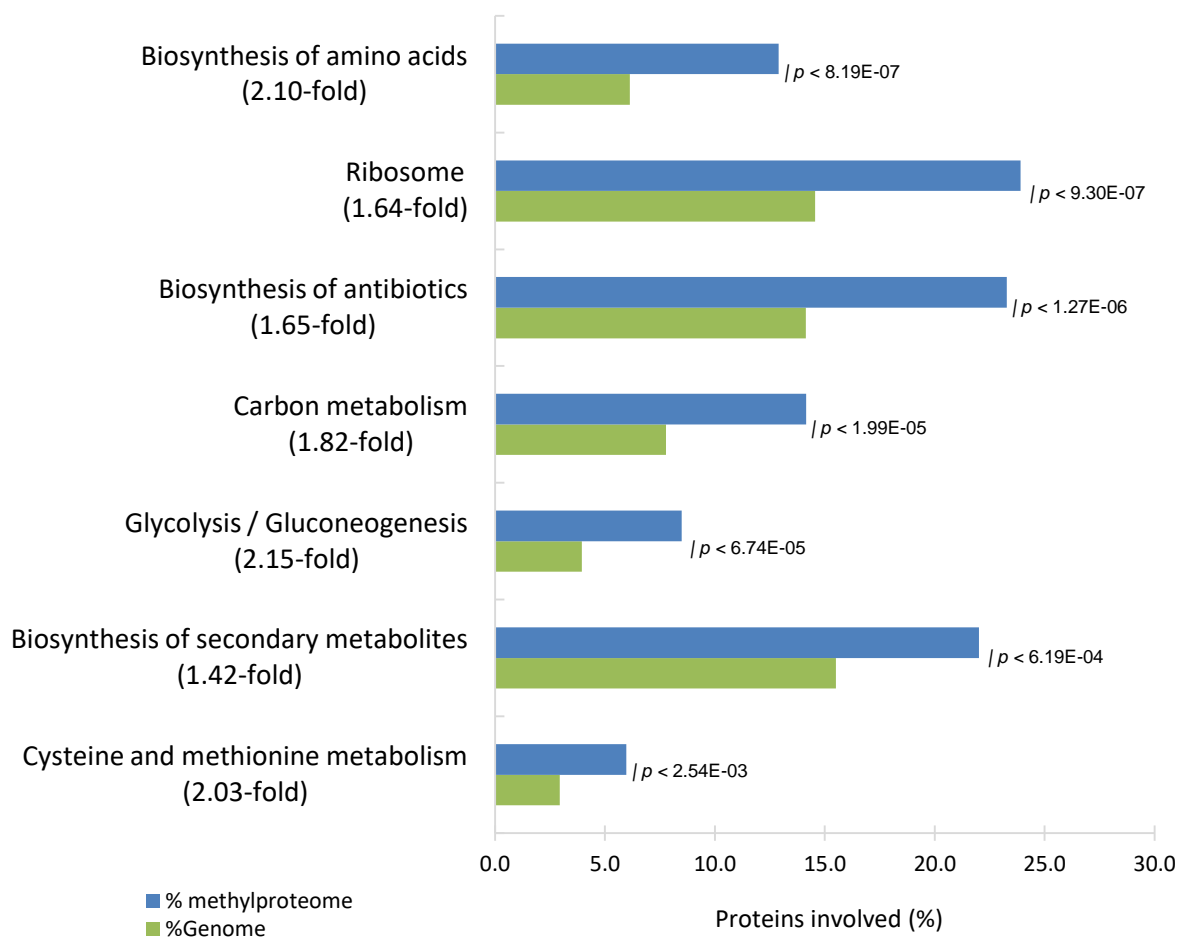
Já a classificação KEGG retornou 203 (23%) proteínas metiladas com termos anotados, distribuídas em 75 diferentes classes (Tabela S3), sendo 30 com 5 ou mais proteínas. Como mostra a Figura 4.16, *metabolic pathways* é a principal classe de KEGG, com 84 proteínas anotadas, seguido de *biosynthesis of secondary metabolites* e *biosynthesis of antibiotics* que juntas somam 91 proteínas anotadas. Interessantemente como quarta classe mais representada está *ribosome* com 38 proteínas e apresenta-se como uma via significativamente enriquecida (Figura 4.17). Além desta, também se destaca o metabolismo de aminoácidos e de açúcares significativamente representado em *T. cruzi*. Ribosomo também aparece como uma via enriquecida.

Figura 4.16 – Proteínas metiladas de *T. cruzi* classificadas de acordo com KEGG.



Fonte: o autor

Figura 4.17 - KEGGs enriquecidas dentre as proteínas metiladas de *T. cruzi*



Fonte: o autor.

A análise de clusterização de alta estrigência das 878 proteínas metiladas, organizou 28 grupos de proteínas, envolvidos com diferentes processos importantes ao *T. cruzi*, dentre eles, processo patogênico, onde 11 *trans-sialidades* foram agrupadas. Outro grupo contém 10 proteínas envolvidas no folding proteico, como chaperonas e chaperoninas (Tabela S5).

#### 4.2.3.2 Panorama das vias e categorias funcionais envolvidas com os metilproteomas de arginina e lisina em epimastigotas de *T. cruzi*

Os estudos de caracterização do metilproteoma geralmente têm analisado separadamente a metilação em arginina (R) e lisina (K) e diferentes processos têm sido identificados como sendo impactados pela modificação nestes diferentes resíduos. Enquanto proteínas metiladas em R estão envolvidas no processamento de RNA, regulação transcricional e reparação de danos no DNA, a metilação em K tem sido amplamente descrita em muitos resíduos de histonas, desempenhando um papel na regulação da compactação da cromatina e da transcrição de genes. Em nosso trabalho caracterizamos a metilação em ambos os resíduos. Desse modo, uma vez identificada a metilação em R e K em *T. cruzi*, foi então avaliado se diferentes funções são reguladas pela metilação nestes dois metilproteomas. Para tanto, foi realizada novamente a análise de enriquecimento, agora com os conjuntos de R e K separadamente (Tabela 4.5).

Como pode ser visto na Tabela 4.5, exceto pela classe Citoplasma em Componente Celular, que aparece tanto para as R metiladas quanto para as K metiladas, as categorias funcionais enriquecidas são diferentes entre a metilação em R e em K. Quando se observa as categorias funcionais mais significativas, a metilação em R impacta diferentes processos como *Oxidation-reduction process* e *Carbohydrate metabolic process* (BP), *Proton-transporting two-sector ATPase complex, catalytic domain* (CC), *Pyridoxal phosphate binding* (MF), ou seja, processos de oxiredução, metabolismo de carboidratos, entre outras vias metabólicas. Por outro lado, a metilação em K revela estar fortemente associada ao processo de tradução, pois aparece enriquecida para *Amino acid transport and metabolism* (COG), *Translation* (BP), *Ribosome* (CC), *Structural constituent of ribosome* (MF) e *Ribosome* (KEGG).

Tabela 4.5 – Categorias funcionais enriquecidas dentre proteínas metiladas em arginina ou lisina em *T. cruzi*.

Category	Residue	Term	Count	Fold Enrichment	PValue	Benjamini
<b>Clusters of Orthologous Groups</b>	<b>K</b>	Amino acid transport and metabolism	11	2.89	2.94E-03	4.89E-02
		Translation	50	2.66	8.93E-11	8.84E-09
<b>Biological Process</b>	<b>K</b>	Cysteine biosynthetic process from serine	8	10.62	1.66E-06	8.23E-05
		Response to stress	7	7.30	1.80E-04	5.92E-03
		Glycolytic process	9	4.69	3.96E-04	9.76E-03
		Oxidation-reduction process	6	8.37	4.00E-04	4.11E-02
	<b>R</b>	Biosynthetic process	10	4.19	4.29E-04	2.23E-02
		Carbohydrate metabolic process	10	3.99	6.31E-04	2.18E-02
		Ribosome	45	4.22	1.04E-16	4.77E-15
<b>Cellular Component</b>	<b>K</b>	Large ribosomal subunit	9	10.97	5.25E-07	1.13E-05
		Cytoplasm	30	2.65	1.97E-06	2.82E-05
		Cell	12	4.41	6.31E-05	6.78E-04
		Endoplasmic reticulum	5	7.24	4.04E-03	3.42E-02
		Proteasome activator complex	3	23.16	5.37E-03	3.79E-02
		<b>R</b>	Cytoplasm	34	3.25	2.00E-09
	Cytosolic small ribosomal subunit		4	14.32	1.92E-03	4.15E-02
	Proton-transporting two-sector ATPase complex, catalytic domain		3	25.06	4.59E-03	4.94E-02
	Structural constituent of ribosome		52	2.97	1.19E-12	1.84E-10
	<b>Molecular Function</b>	<b>K</b>	GTPase activity	16	4.00	6.02E-06
Cystathionine beta-synthase activity			6	13.73	1.73E-05	8.94E-04
Translation elongation factor activity			16	3.16	1.16E-04	4.47E-03
NADP binding			7	7.01	2.64E-04	8.14E-03
Pyridoxal phosphate binding			12	6.21	1.72E-06	2.55E-04
<b>R</b>		L-alanine:2-oxoglutarate aminotransferase	4	20.19	4.59E-04	3.34E-02
		Ribosome	56	1.87	6.17E-07	4.01E-05
<b>KEGG Pathways</b>	<b>K</b>	Biosynthesis of amino acids	26	2.06	3.41E-04	1.10E-02
		Biosynthesis of antibiotics	47	1.61	3.99E-04	8.61E-03
		Cysteine and methionine metabolism	16	2.64	4.80E-04	7.76E-03
		Glycolysis / Gluconeogenesis	19	2.34	5.92E-04	7.66E-03
		Glycine, serine and threonine metabolism	11	2.75	3.98E-03	4.23E-02

Fonte: o autor.

As categorias foram filtradas com cutoff de Benjamini-Hochberg  $\leq 0.05$  e as que passaram o cutoff ordenadas por valor crescente de *p*value.

Em seguida, para garantir uma comparação justa de nossos dados com os disponíveis na literatura, os trabalhos recentes que caracterizaram o metilproteoma de R em *T. brucei* (FISK et al., 2013; LOTT et al., 2013), únicos disponíveis até então para tripanossomatídeos, foram reanalisados com as mesmas ferramentas e parâmetros utilizados para *T. cruzi* no presente trabalho. Conforme esperado, para a forma procíclica do parasito africano, as proteínas metiladas em arginina estão associadas ao citoesqueleto e envolvidas com locomoção, incluindo a presença de cinesinas (Tabela 4.6). A comparação entre os processos impactados pelo metilproteoma de arginina em *T. cruzi* e *T. brucei* apresenta diferenças marcantes.

Tabela 4.6 - Categorias funcionais enriquecidas no conjunto de proteínas metiladas em arginina em *T. brucei*.

Category	Residue	Term	Count	Fold Enrichment	PValue	Benjamini
<b>Biological Process</b>	R	Microtubule-based movement	25	3.89	3.55E-09	1E-06
<b>Molecular Function</b>	R	Microtubule motor activity	22	3.60	1.53E-07	3.35E-05
<b>Molecular Function</b>	R	ATPase activity	22	3.24	1.17E-06	0.000128
<b>Cellular Component</b>	R	Kinesin complex	14	3.77	4.05E-05	0.005086
<b>Molecular Function</b>	R	ATP binding	89	1.46	5.77E-05	0.004204

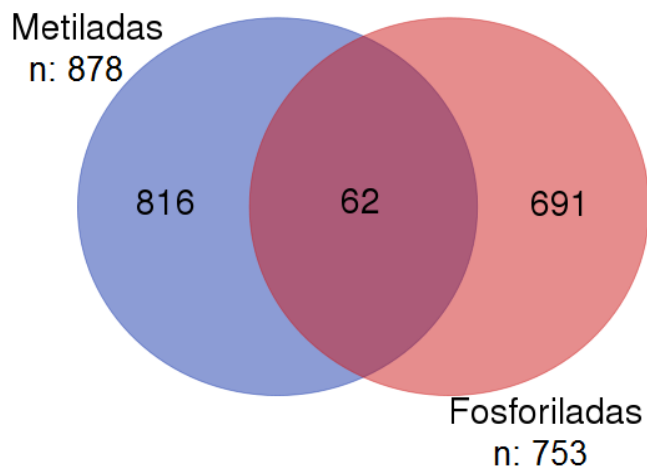
Fonte: o autor.

#### 4.2.3.3 Metilação apresenta coocorrência com fosforilação em proteínas de *T. cruzi*

Uma vez que diferentes modificações pós-traducionais (PTMs) ocorrem em uma mesma proteína, isso permite uma ação sinérgica ou antagônica sobre esta, pois dependendo de sua combinação, o *crosstalk* entre estas diferentes modificações desencadeia outras funções na proteína, afetando suas interações com outros componentes celulares, por exemplo. Neste sentido, a fim de verificar a coocorrência de metilação e fosforilação na mesma proteína em *T. cruzi* foi realizada a busca por proteínas concomitantemente metiladas e fosforiladas, a partir da reanálise dos dados já reportados no grupo (MARCHINI et al., 2011).

Das 878 proteínas identificadas com sítios de metilação (Me, Me<sub>2</sub> ou Me<sub>3</sub>), 62 estavam também fosforiladas dentro do conjunto das 753 fosforiladas identificadas por Marchini e colaboradores (Figura 4.18). Dentre as 62 proteínas onde coocorrem sítios de metilação e fosforilação, 35 são anotadas como hipotéticas, dentre as quais 33 como hipotéticas conservadas e duas como hipotéticas não conservadas (Tabela 4.7).

Figura 4.18 – Sobreposição do metilproteoma e fosfoproteoma em *T. cruzi*.



Fonte: o autor

Em azul o conjunto de proteínas metiladas identificadas neste trabalho. Em rosa o conjunto das proteínas fosforiladas identificadas por Marchini e colaboradores (2011).

Tabela 4.7 - Proteínas que apresentam coocorrência de metilação e fosforilação em *T. cruzi*.

Descrição	Uniprot Ids	Gene Ids	Posição sítio Metilação (K,R)	Posição sítio Fosforilação (S,T,Y)
beta tubulin, putative	Q4DQP2	Tc00.1047053506563.40	R2; K19; R262; K297; R318; K350; R379; R390	S285
calpain cysteine peptidase, putative	Q4CPQ6	Tc00.1047053505985.9	R1356	S334; T335; T337; S339; S346
calpain-like cysteine peptidase, clan CA, family C2, putative	Q4CW64	Tc00.1047053509003.30	K72	S20
cysteine peptidase, Clan CA, family C2, putative	Q4D6J0	Tc00.1047053506721.30	R3348; R3357	S1889; T4555
cysteine peptidase, Clan CA, family C2, putative	Q4E0D8 Q4CL00	Tc00.1047053508999.190 Tc00.1047053510957.9	K732	T812; Y818; S822
cAMP-specific phosphodiesterase, putative	Q4DUD4	Tc00.1047053506625.80	R635; R636	S485; S489
cytoskeleton-associated protein CAP5.5, putative	Q4DJS7	Tc00.1047053509237.130	R214	S759; S769
dynein heavy chain, putative	Q4DX29 Q4CPP9	Tc00.1047053508831.4 Tc00.1047053508275.9	K124; R126; K131	S189; S483
eukaryotic initiation factor 5a, putative	Q4E4N4	Tc00.1047053506925.120 Tc00.1047053506925.130	K42	S2
heterogeneous nuclear ribonucleoprotein H/F, putative	Q4DGN8 Q4DFM4	Tc00.1047053504157.10 Tc00.1047053511109.130	K147	S251; S339
paraflagellar rod protein, putative	Q4DNY3	Tc00.1047053509099.30	K564	S570; T571; S574; S642
phosphoprotein phosphatase, putative	Q4DDM9	Tc00.1047053507601.10	K625	S5; S8; S9; T232; S270; T272
prostaglandin F2alpha synthase	Q4E4V7 Q4CNI1	Tc00.1047053508461.80 Tc00.1047053507617.9	R9	S215; S243
protein kinase A catalytic subunit isoform 2, putative	Q4E4T3	Tc00.1047053508461.280	K18; K160	T174
protein kinase, putative	Q4D5A8	Tc00.1047053511633.70	R387; R389	S318; S319; S322
pyruvate dehydrogenase E1 component alpha subunit, putative	Q4DI66	Tc00.1047053507831.70	R59	S284
pyruvate phosphate dikinase, putative	Q4E0Q0 Q4E3P5	Tc00.1047053506297.190 Tc00.1047053510101.140	R365; R587; K871	T481; S482; S680
retrotransposon hot spot (RHS) protein, putative	Q4E3H4	Tc00.1047053507611.10	K553	S32; S34; T36
RNA guanylyltransferase, putative	Q4D3H7	Tc00.1047053507511.30	K445	S30; S34; S35; S37
RNA-binding protein, putative	Q4E5D6	Tc00.1047053511277.580	R8; R251	S302; S304; S307; S308; S412
RNA-binding protein, putative	Q4D7Y1 Q4D3A9	Tc00.1047053510755.120 Tc00.1047053508413.50	R213	S6; S7; T8; S9; S175; S182 ; S188

<b>RNA-binding protein, putative</b>	Q4D488	Tc00.1047053509317.60	R36	S375;S377;S379;S409;
	Q4DT61	Tc00.1047053511621.50		S411;S413
<b>sterol 24-c-methyltransferase, putative</b>	Q4CLW8	Tc00.1047053505683.10	R91	S235
	Q4CMB7	Tc00.1047053504191.10		
	Q4CM63	Tc00.1047053510185.10		
<b>sterol 24-c-methyltransferase, putative</b>	Q4CMB7	Tc00.1047053504191.10	R91	S235
	Q4CM63	Tc00.1047053510185.10		
<b>tyrosine aminotransferase</b>	Q4CVI0	Tc00.1047053510795.10	K60	S77
<b>tyrosine aminotransferase</b>	Q4E4E7	Tc00.1047053510187.30	R196	S77
		Tc00.1047053510187.40		
<b>ubiquitin hydrolase, putative</b>	Q4DG39	Tc00.1047053510761.70	R540;K544	S947;S949;S952;T955; S956
<b>hypothetical protein</b>	Q4DWC8	Tc00.1047053509733.60	R8	S10;S13;S66;S129; S142;S144;Y146;T183; S185;S248;S402;S403; S411
<b>hypothetical protein</b>	Q4DZM3	Tc00.1047053504147.224	K362;K375	S188;S189;S193;S194
<b>hypothetical protein, conserved</b>	Q4CYL3	Tc00.1047053504423.30	K308;R408	S57;S618;T622;S813
<b>hypothetical protein, conserved</b>	Q4D0U3	Tc00.1047053508051.20	R1419	S256;S260;S265;S266; S267;S944;S947;S948; S1403;S1406
<b>hypothetical protein, conserved</b>	Q4CYE7	Tc00.1047053507629.30	R938;R946	S762;S764;S767
<b>hypothetical protein, conserved</b>	Q4DRF1	Tc00.1047053508547.160	K756;K760	S15;S81;T83;S86;S619; S620
<b>hypothetical protein, conserved</b>	Q4DVF8	Tc00.1047053506859.180	R157;K209	S12;S14;T21
<b>hypothetical protein, conserved</b>	Q4DIU6	Tc00.1047053511389.50	R237;R239	T221
<b>hypothetical protein, conserved</b>	Q4DPA5	Tc00.1047053509647.130	K7;R9	S91;S95
<b>hypothetical protein, conserved</b>	Q4D770	Tc00.1047053508265.100	R231;K364	S91;T114;S274;S653; S695;S987;S1005; S1132;T1137;S1139; S1150;S1153;S1154; S1159;T1164
<b>hypothetical protein, conserved</b>	Q4E4Q4	Tc00.1047053508461.550	R16	S315;S316
<b>hypothetical protein, conserved</b>	Q4DNH8	Tc00.1047053509791.189	R146	S77;S79
<b>hypothetical protein, conserved</b>	Q4CTC8	Tc00.1047053507569.10	R266;K308;R408	S57;S241;T246;S618; T622;S810
<b>hypothetical protein, conserved</b>	Q4DG58	Tc00.1047053511693.20	R200	S90;S92
<b>hypothetical protein, conserved</b>	Q4D3F7	Tc00.1047053508719.70	K146	S77;S79;S230;S276; S332;S344;S381



hypothetical protein, conserved	Q4CSI2	Tc00.1047053506441.20	R231;K364	S91;T114;S274;S615; Y618;S619;S653;S674; S892;S927;T1053; S1075;S1078;T1080
hypothetical protein, conserved	Q4CP52	Tc00.1047053503539.20	R4;R7	S362;S363;S365;S429; S436;S441;S442;S443; S444;S445;Y447
hypothetical protein, conserved	Q4DHI4	Tc00.1047053508569.90	R298	S121;S123;S125
hypothetical protein, conserved	Q4D6H3	Tc00.1047053509247.30	R794;K800;K804	S825
hypothetical protein, conserved	Q4D015	Tc00.1047053511803.40	K89	S502;T505
hypothetical protein, conserved	Q4E2Q5	Tc00.1047053504153.280	R171	S366;S370
hypothetical protein, conserved	Q4D2L5	Tc00.1047053511435.40	R115;R118;R126	S70;S88;T90;S91;S305; S311;S352;S355;S359
hypothetical protein, conserved	Q4D442	Tc00.1047053510733.50	K70;K77	S484
hypothetical protein, conserved	Q4D0P8	Tc00.1047053510301.40	K305;K313	S97;S99;T105
	Q4D4R7	Tc00.1047053504163.70		
hypothetical protein, conserved	Q4CWG6	Tc00.1047053508145.49	R4;R8	S1788;S1792;S1793; S1796
	Q4D106	Tc00.1047053504243.30		
hypothetical protein, conserved	Q4DZJ0	Tc00.1047053508799.270	R412	T588
	Q4CMC7	Tc00.1047053506551.10	R412	
	Q4DZ19	Tc00.1047053508799.280	R401	
	Q4CUG0	Tc00.1047053509713.10		
hypothetical protein, conserved	Q4CVJ1	Tc00.1047053504199.20	R96	S128;S130;S132
	Q4DIP8	Tc00.1047053506635.130		
hypothetical protein, conserved	Q4DJ92	Tc00.1047053511815.170	R322;K374	S427;S428
	Q4DVF9	Tc00.1047053506859.170		
hypothetical protein, conserved	Q4CY87	Tc00.1047053510001.20	R507;K508	S17;S21;S44;S48;S51; S56;S491;T494;S497; S502
	Q4D238	Tc00.1047053507529.20		
hypothetical protein, conserved	Q4CYY8	Tc00.1047053509877.60	K143	S65
	Q4DRV7	Tc00.1047053507669.139		
hypothetical protein, conserved	Q4E5E4	Tc00.1047053511277.490	R7;R10	S186;S189
	Q4E1X3	Tc00.1047053507993.40		
hypothetical protein, conserved	Q4E260	Tc00.1047053508707.170	R174;K176;K177	S255;T256;T259;T646
	Q4DCS6	Tc00.1047053509073.40		
hypothetical protein, conserved	Q4D9L9	Tc00.1047053510579.50	K478;K485;R492	S251;T254;S255;T256; S257;S258;S404;S407; S597;S599
	Q4DD47	Tc00.1047053510241.90		
hypothetical protein, conserved	Q4CZT7	Tc00.1047053506151.20	K1376;K1377	S114
	Q4D7V8	Tc00.1047053511407.50		
hypothetical protein, conserved	Q4E246	Tc00.1047053508707.310	R512;K548;K632	S898;S906;S910

Fonte: o autor.

Para melhor avaliar a quais vias e processos estão envolvidas as proteínas que apresentam coocorrência de metilação e fosforilação, foi realizada a análise de clusterização funcional através da ferramenta DAVID de Bioinformática, utilizando parâmetros de alta estrigência para associação de termos, conforme definidos pela plataforma. A partir das anotações nos bancos INTERPRO, SMART, GO e KEGG pathways, a análise identificou a presença de três clusters funcionais (Tabela 4.8). A classificação funcional completa das 62 proteínas metiladas e fosforiladas está disponível na Tabela Suplementar S6.

Tabela 4.8 - Clusterização funcional das proteínas que apresentam coocorrência de metilação e fosforilação em *T. cruzi*.

Annotation Cluster 1		Enrichment Score: 2.75				
Category	Term ID	Description	Count	%	PValue	Benjamini
INTERPRO	IPR012677	Nucleotide-binding, alpha-beta plait	6	9.68	2.68E-04	2.36E-02
GOTERM_MF	GO:0000166	nucleotide binding	6	9.68	7.65E-04	2.34E-02
INTERPRO	IPR000504	RNA recognition motif domain	5	8.06	9.90E-04	4.31E-02
SMART	SM00360	RRM	5	8.06	2.34E-03	3.22E-02
GOTERM_MF	GO:0003676	nucleic acid binding	5	8.06	3.80E-02	3.30E-01

Annotation Cluster 2		Enrichment Score: 2.20				
Category	Term ID	Description	Count	%	PValue	Benjamini
INTERPRO	IPR022684	Peptidase C2, calpain family	3	4.84	4.12E-03	1.15E-01
GOTERM_CC	GO:0005622	intracellular	4	6.45	5.61E-03	4.40E-02
INTERPRO	IPR001300	Peptidase C2, calpain, catalytic domain	3	4.84	5.75E-03	9.76E-02
GOTERM_MF	GO:0004198	calcium-dependent cysteine-type endopeptidase activity	3	4.84	6.89E-03	1.02E-01
SMART	SM00230	CysPc	3	4.84	1.04E-02	7.08E-02

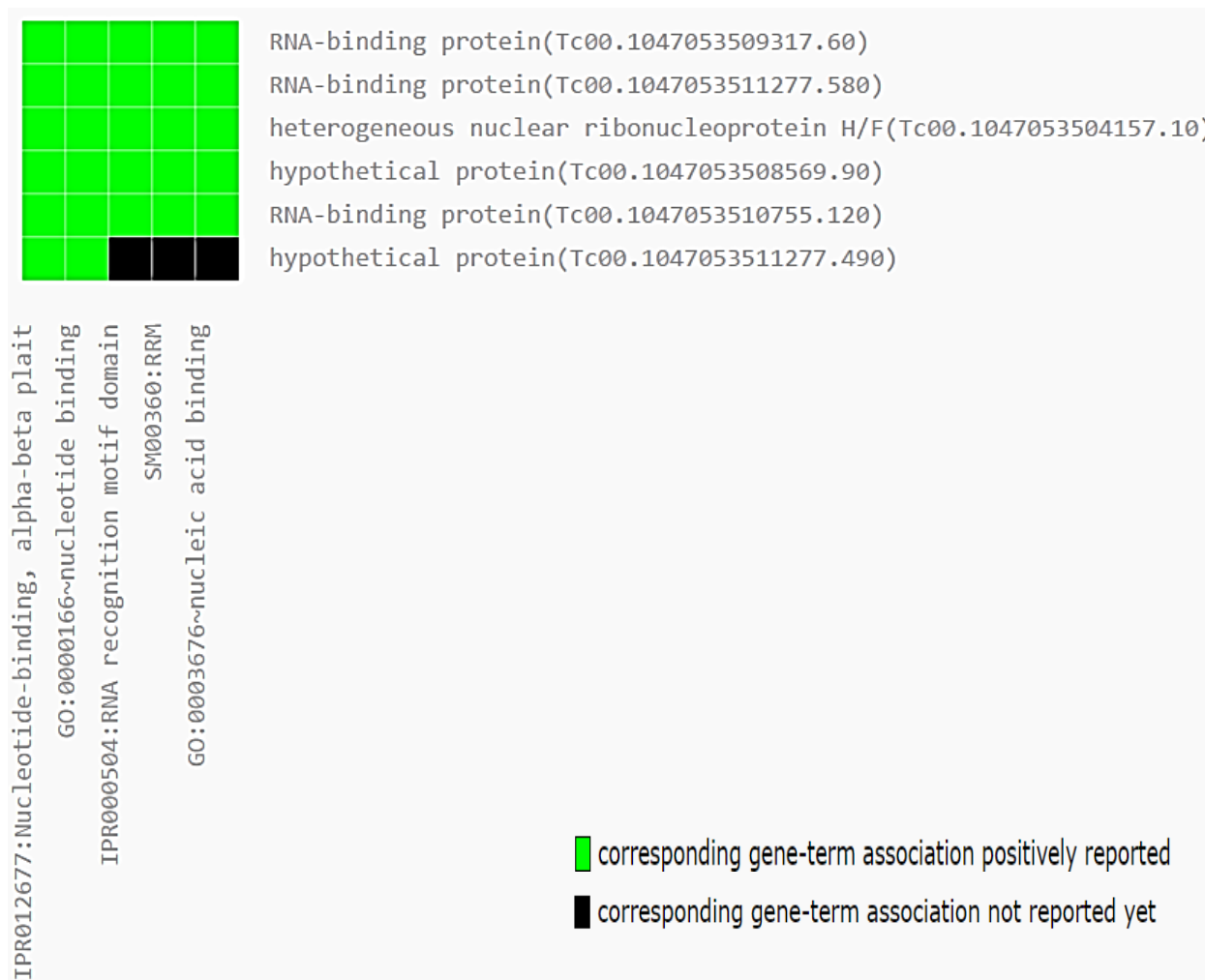
  

Annotation Cluster 3		Enrichment Score: 1.66				
Category	Term ID	Description	Count	%	PValue	Benjamini
KEGG_PATHWAY	tcr01130	Biosynthesis of antibiotics	5	8.06	9.63E-03	1.43E-01
KEGG_PATHWAY	tcr01110	Biosynthesis of secondary metabolites	5	8.06	1.34E-02	1.02E-01
KEGG_PATHWAY	tcr01100	Metabolic pathways	6	9.68	8.34E-02	3.71E-01

Fonte: o autor.

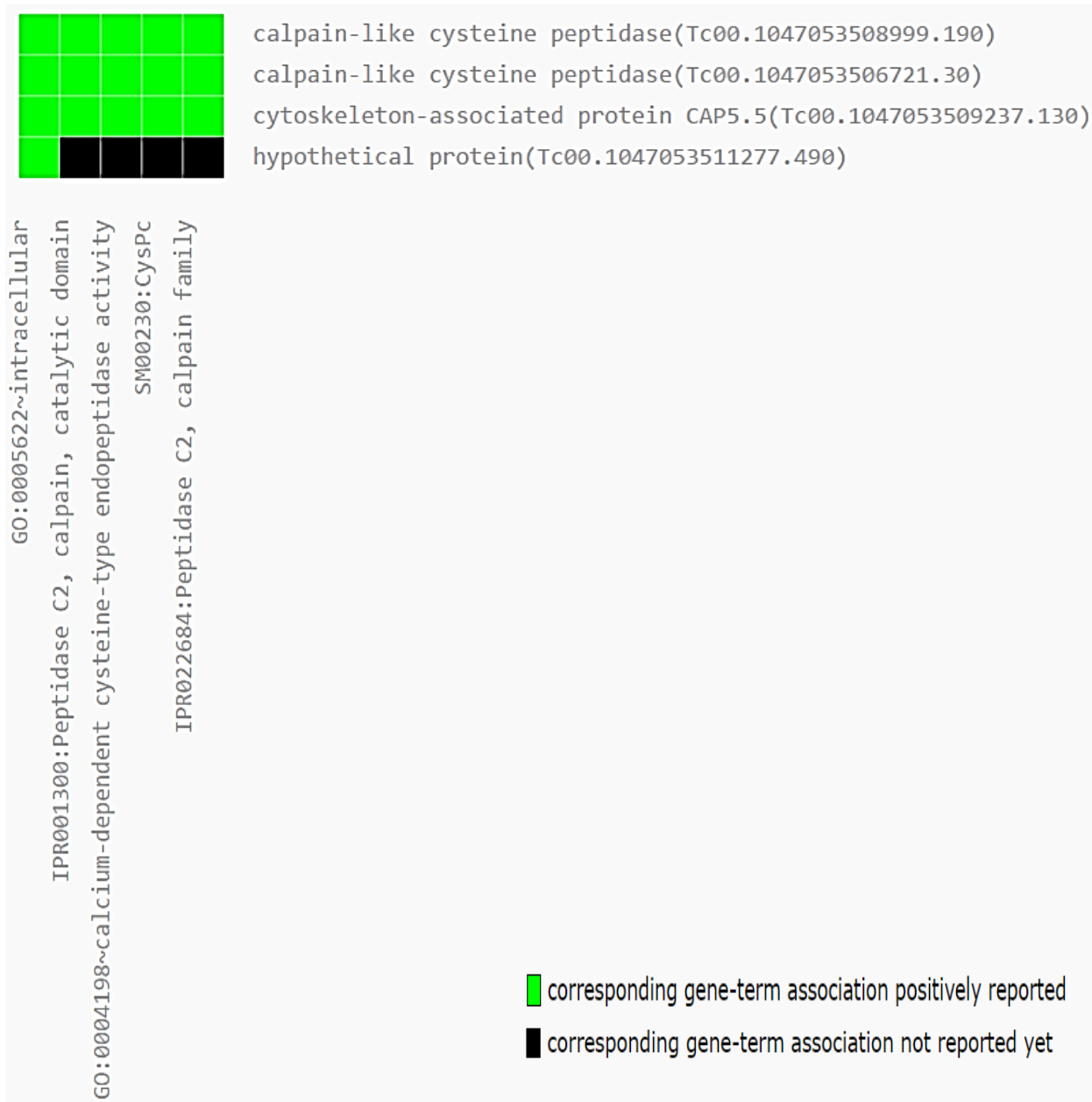
No primeiro cluster há 5 proteínas que estão envolvidas com a ligação a RNA e pertencem a família das proteínas com motivos de reconhecimento a RNA (RRM) (Figura 4.19). O segundo cluster agrupa 4 proteínas com atividade peptidases de cisteína, da família *Peptidase C2, calpain*, dentre as quais está a calpaína, que uma protease intracelular envolvida em muitas funções celulares que são reguladas por cálcio (Figura 4.20). O terceiro cluster representado, contendo 6 proteínas, foi associado principalmente às vias que envolvem a biossíntese de metabólitos secundários e antibióticos (Figura 4.21).

Figura 4.19 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em *T. cruzi* e categorias funcionais do Cluster de Anotações 1.



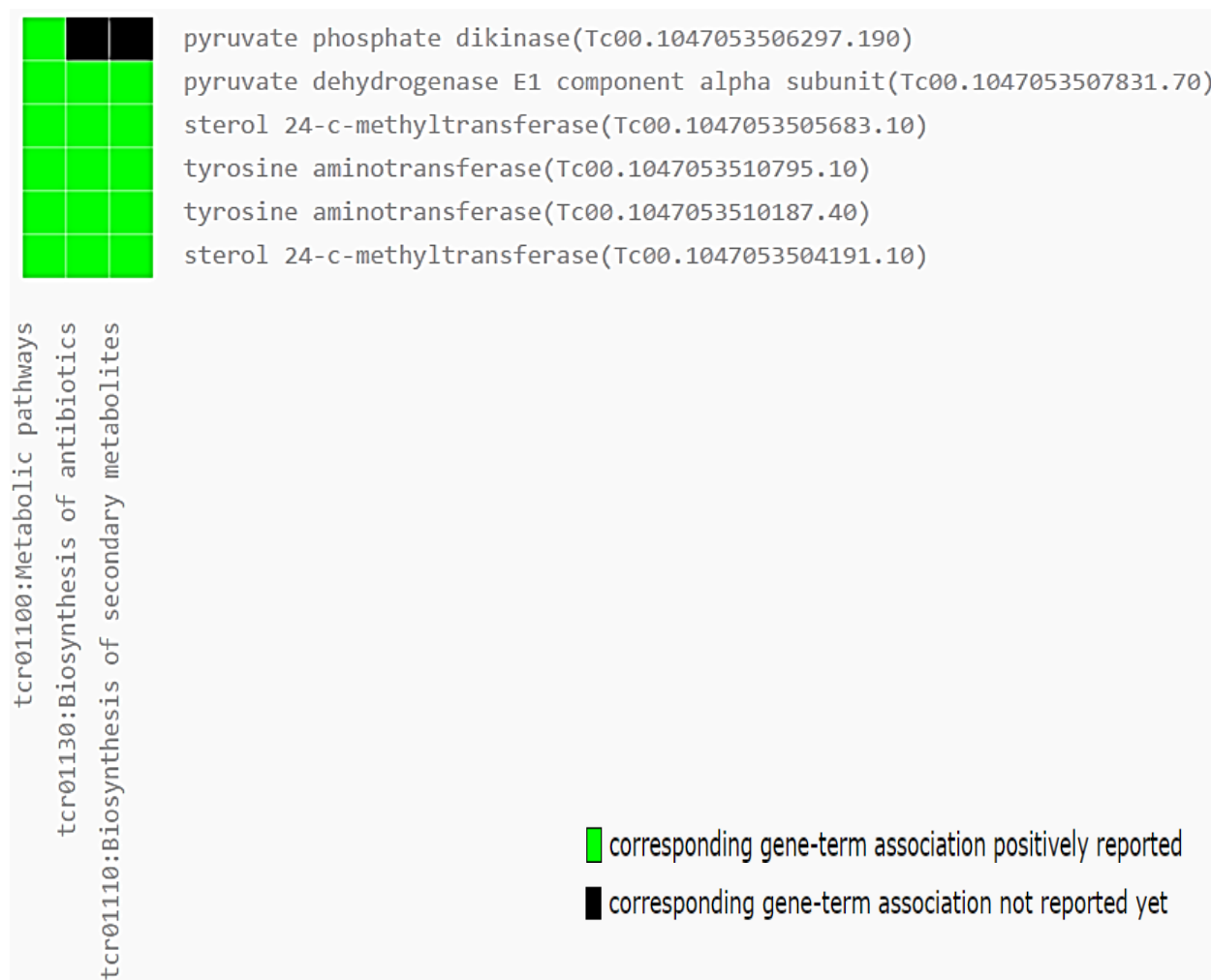
Fonte: o autor.

Figura 4.20 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em *T. cruzi* e categorias funcionais do Cluster de Anotações 2.



Fonte: o autor.

Figura 4.21 – Associação entre proteínas que apresentam coocorrência de metilação e fosforilação em *T. cruzi* e categorias funcionais do Cluster de Anotações 3.



Fonte: o autor.

## 5 DISCUSSÃO

### 5.1 *T. cruzi* possui uma maquinaria de metilação e demetilação conservada

A metilação pós-traducional de resíduos de arginina (R) e de lisina (K) é realizada por famílias de enzimas PRMTs e PKMTs, respectivamente. A fim de verificar a existência da maquinaria de metilação em *T. cruzi*, isto é, a presença de PRMTs e PKMTs neste parasito, condição *sine qua non* para o processo de metilação de proteínas, foi realizada a busca destas proteínas em humanos e *T. brucei* já reportados pela literatura e nas informações disponíveis nos bancos de dados, assim como verificada a presença de domínios de metiltransferase e demetilases em suas sequências. A busca por domínios é uma interessante estratégia para inferir funções à proteínas hipotéticas e sem nenhuma caracterização como as proteínas de *T. cruzi* identificadas neste trabalho. Para esta finalidade o emprego do CDD permite a identificação de domínios conservados através da sequência proteica usando uma sequência *query* para encontrar os domínios em um banco de dados, baseado no *score* de uma matriz de alinhamento que utiliza como fonte os dados disponíveis nos bancos Pfam (<http://pfam.xfam.org>), SMART (<http://smart.embl.de>), COGs (<http://www.ncbi.nlm.nih.gov/COG>), TIGRFAMs (<http://www.jcvi.org/tigrfams>) e NCBI *Protein Clusters* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters>) (MARCHLER-BAUER et al., 2015).

Em humanos, são conhecidas nove PRMTs (MORALES et al., 2016). Já para *T. brucei*, cinco PRMTs são descritas até o momento (FISK; READ, 2011). Destas, TbPRMT1, TbPRMT3 e TbPRMT5 foram previamente consideradas homólogas às de humanos (FIGUEIREDO; CROSS; JANZEN, 2009) e, no presente trabalho, demonstrou-se que o mesmo é verdadeiro para TbPRMT6 e TbPRMT7, conforme alinhamento (Figura 4.3). Do mesmo modo, estas cinco TbPRMTs possuem homólogas também em *T. cruzi* (Figura 4.3).

Mesmo este parasito carecendo de trabalhos de caracterização, a homologia destas cinco TcPRMTs com *T. brucei* e humanos, aliada à presença de domínios de metiltransferases em suas sequências (Figura 4.2), sugerem a presença de proteínas com a função de metiltransferases de arginina em *T. cruzi*.

Dois tipos de PKMTs são conhecidos em humanos e em tripanossomas, os quais podem mono, di ou trimetilar resíduos de lisina: as com domínios SET e os homólogos da família DOT1. Da família SET há mais de vinte genes codificadores no genoma de *T. brucei*, no entanto, nenhum ainda foi caracterizado. As enzimas da família DOT1 possuem dois homólogos descritos em *T. brucei*, DOT1A e DOT1B (Tabela 4.2), as quais estão envolvidas com a dimetilação e trimetilação da lisina 76 na histona H3 (H3K76), respectivamente (FIGUEIREDO; CROSS; JANZEN, 2009). Já em humanos, DOT1L (também conhecida como PKMT4), tem a função bem conhecida de mono, di ou trimetilar especificamente a lisina 79 da histona H3 (H3K79) (FAROOQ et al., 2016). O resíduo 76 de lisina em *T. brucei* e o resíduo 79 de lisina em humanos, são, provavelmente, sítios funcionalmente equivalentes em suas respectivas histonas. Em nosso trabalho, a histona H3 foi identificada trimetilada no resíduo de lisina 77 (H3K77) em *T. cruzi* no fracionamento a partir de gel gradiente SDS-PAGE.

Apresentamos também a análise das demetilases contendo o domínio JmjC. A análise bioinformática e a pesquisa da literatura, forneceu a homologia do domínio JmjC entre *T. cruzi*, *T. brucei* e humanos, a julgar pela presença do domínio e inferência de sua possível função. As proteínas reportadas em *T. cruzi* possuem este domínio e satisfazem assim os requerimentos de ligação aos cofatores Fe (II) e  $\alpha$ -cetoglutarato parecendo ótimas candidatas a demetilases de histonas e mesmo de proteínas não-histonas conforme Klose, Kallin e Zhang (2006). Dessa forma, essas proteínas de *T. cruzi* (Tabela 4.3) foram associadas à função de demetilase, mesmo até o momento estarem descritas como proteínas hipotéticas conservadas neste organismo, provavelmente estas proteínas funcionem como proteínas hidroxilases ou histona demetilases, exercendo em *T. cruzi* o antagonismo da metilação em lisinas. Além disso, possivelmente TcJmjD6 (TcCLB.508355.310) em *T. cruzi* tenha papel bifuncional como demetilase da mesma forma que sua homóloga HsJmjD6 (NP\_055982.2) em humanos e, desse modo, esta demetilase também seja responsável por demetilar resíduos de arginina.

Os alinhamentos das PRMTS/PKMTs e das demetilases (JmjC), juntamente com a determinação dos domínios em suas sequências de ligação à molécula AdoMet e JmjC que associam estas proteínas à família das metiltransferases e demetilases, respectivamente, revelam que *T. cruzi* apresenta uma potencial maquinaria de metilação, e assim como *T.*

*brucei* e humanos, apresenta PRMTs, PKMTs e demetilases (JmjC) com domínios catalíticos conservados (Figuras 4.2 e 4.4). As sequências de *T. cruzi* apresentaram domínios com uma alta associação com metiltransferase e demetilases, assim como as sequências de *T. brucei* e humanos, nos quais já era esperado conforme o que foi reportado na literatura.

## **5.2 A otimização do preparo de amostras permitiu a identificação em larga escala dos sítios de metilação encontrados em *T. cruzi***

Após a identificação da presença da maquinaria de metilação em *T. cruzi*, foi realizada a busca preliminar das proteínas metiladas em *T. cruzi*, a partir de dados proteômicos já disponíveis. Esta análise demonstrou, pela primeira vez, que esta modificação está presente neste organismo de maneira mais ampla, e está relacionada a uma série de funções em diferentes locais da célula, como esperado, devido à importância dessa modificação em organismos eucariotos e procariotos.

A fim de ampliar a identificação de proteínas metiladas, para melhor caracterizar o metilproteoma de *T. cruzi* e verificar o quão abrangente é sua ocorrência neste parasito, foram buscados protocolos na literatura utilizados para a identificação de proteínas metiladas em larga escala por proteômica LC-MS/MS em humanos e tripanossomatídeos (FISK et al., 2013; LOTT et al., 2013; RAPPSILBER; MANN; ISHIHAMA, 2007; SYLVESTERSEN et al., 2014; SYLVESTERSEN; NIELSEN, 2015; UHLMANN et al., 2012), os quais foram testados para verificar sua eficiência na identificação de proteínas metiladas em *T. cruzi*.

Inicialmente foi buscado melhorar a extração de proteínas das células, para isso foi empregada o protocolo FASP, onde a lise na presença de detergente (SDS) permite alta solubilização das proteínas, inclusive proteínas de membrana. O DTT permite a redução das pontes dissulfeto (-S-S-) nas proteínas e ureia 8M desta etapa desnatura fortemente as proteínas e dissocia as micelas e agregados proteicos formados durante a lise celular., ao passo que a iodoacetamida é um agente alquilante, que neste caso carboxiamidometila os tióis das cisteínas nas proteínas fazendo com que não refaçam as pontes dissulfeto (-S-S-). Outra vantagem é que esse dispositivo de filtração permite, através das centrifugações, a remoção do SDS utilizado na lise celular e a ureia da desnaturação proteica, o DTT e componentes de baixo peso molecular permitindo melhor digestão e posterior eluição de



peptídeos mais puros para análise por LC-MS/MS. Após a digestão, as lavagens com ABC reduzem a quantidade de Ureia 8M remanescente da etapa anterior para a digestão com tripsina (WIŚNIEWSKI et al., 2009).

Como intuito de otimizar a análise dos peptídeos, reduzindo a quantidade de amostra e materiais necessários no seu fracionamento e armazenamento, foram utilizadas microcolunas *StageTips*. Estas microcolunas são confeccionadas *in-house* a partir na colocação de discos 3M Empore™ que possuem diferentes tipos de *beads* cromatográficas (ex.: C18, C8, SCX, carvão ativado, etc.) imobilizadas em uma malha de Teflon em ponteiras (*tips*) de variados tamanhos (ex.: 1000 µL, 200 µL, 10 µL), de acordo com a capacidade de carregamento desejada. Apresentam versatilidade pois porções adicionais de disco podem ser colocadas na tip, assim como combinar discos com diferentes funcionalidades cromatográficas. As *tips* comerciais para essa finalidade, chamadas ZipTips (Millipore), possuem custos comparativamente elevados, têm limitações para o trabalho de alta sensibilidade devido a problemas de recuperação e não podem ser ajustadas de acordo com a necessidade de carregamento de amostra. Além disso, a sua funcionalidade não pode ser expandida ou combinadas *in-house* como nas *StageTip* aqui descritas. Estas combinam, portanto, a flexibilidade e baixo preço com a facilidade de utilização, recuperação, reprodutibilidade e robustez (RAPPSILBER; MANN; ISHIHAMA, 2007).

Uma vez que o fracionamento traz vantagens para a identificação de um número maior de proteínas e sítios, o fracionamento por C18-SCX (sal) em *StageTips*, é um fracionamento relativamente de baixo custo e que demonstrou ser eficiente na identificação de novas proteínas em relação à uma amostra não fracionada, podendo ser aplicado a um experimento quantitativo, pois pode-se ter mais réplicas sem utilizar tanto material.

O protocolo de C18-básica, o qual teve frações concatenadas se mostrou um eficiente protocolo comparativamente à C18-SCX, no entanto envolve mais etapas de fracionamento e manipulação de amostra.

No protocolo de Gel SDS-PAGE, a separação das proteínas por tamanho molecular na eletroforese já pré-fraciona diferentes grupos de proteínas com base em seu peso molecular e o fracionamento de determinadas faixas de tamanho, seguida da digestão com tripsina permitiu uma eficiente separação do conjunto de peptídeos identificados (20577), isso se deve ao fato de que o fracionamento, ao reduzir a complexidade da amostra, permite

que cada fração analisada por LC-MS/MS, tenha um *dynamic range* maior, o que permite nas frações a identificação das proteínas menos abundantes em relação à amostra não fracionada. É importante ressaltar o cuidado com a preparação de amostra empregado na coloração do gel SDS-PAGE, alguns corantes utilizados possuem metanol na composição, o que pode gerar artefatos de metilação em cisteínas, ácidos aspártico e glutâmico, como já reportado em alguns trabalhos (CHEN et al., 2010; NIELSEN; SAVITSKI; ZUBAREV, 2006). Mesmo que nossa busca por metilação tenha sido feita em resíduos de arginina e lisina, para garantir a ideal preparação da amostra, foi utilizado na coloração do gel SDS-PAGE o Coomassie Brilliant Blue G-250 (LAWRENCE; BESIR, 2009), corante cuja composição é livre de solventes orgânicos.

Assim, para o panorama geral de epi (não quantitativo), foi utilizado o experimento de SDS-PAGE, enquanto que na análise quantitativa, a estratégia de C18-SCX é mais atraente, pois o número de proteínas metiladas identificadas neste protocolo é quase tão alto quanto o protocolo por SDS-PAGE, com apenas 4 frações. Além de ter se mostrado eficiente, é mais viável fazer réplicas biológicas a partir deste protocolo para uma análise quantitativa. Fazer réplicas das 10 frações do protocolo SDS-PAGE, talvez aumente ainda mais o número de proteínas identificadas, mas seria inviável economicamente e pelo tempo que demandaria de análise LC-MS/MS.

O aumento no número de proteínas/sítios metilados nos fracionamentos foi devido ao aumento do número geral de identificações, e não a um enriquecimento. Não foi verificado um aumento no número de proteínas metiladas identificadas em uma fração específica. A fim de determinar a conformidade do números de cada um dos protocolos, novos testes comparativos devem ser realizados para averiguar o acréscimo de identificações caso sejam feitas réplicas, e assim verificar a eficiência máxima de cada um dos protocolos.

### **5.3 A classificação funcional das proteínas metiladas sugere uma potencial influência da metilação em diversos processos biológicos de *T. cruzi***

A identificação das 878 proteínas metiladas na forma epimastigota de *T. cruzi*, (Tabela S1), e dos 1336 sítios de metilação, tanto da forma monometilada, quanto nas formas

dimetilada e trimetilada (Me/Me<sub>2</sub>/Me<sub>3</sub>) (Tabela S2) já indicam a significativa presença dessa modificação neste parasito. O passo seguinte à verificação de quais são as proteínas que encontram-se metiladas na forma epimastigota de *T. cruzi*, foi identificar a que classes funcionais e vias estão envolvidas.

No estudo de Lott e colaboradores (2013), foram identificadas as proteínas metiladas em arginina na forma procíclica de *T. brucei*, já a metilação em lisinas não foi abordada. Em nosso trabalho, proteínas modificadas tanto em lisina quanto em arginina foram identificadas, sendo que as principais funções representadas pelas proteínas metiladas na forma epimastigota do *T. cruzi* foram relacionadas às atividades de síntese de proteínas, uma vez que a análise funcional revelou que as proteínas metiladas em epimastigota de *T. cruzi* estão envolvidas com ribossomos (Figura 4.10 e 4.17) e com o processo de tradução (Figura 4.12). No entanto, quando apenas as metilproteínas modificadas em arginina ou em lisina foram analisadas, funções distintas são mais representadas: enquanto as metilproteínas de lisina estão associadas, principalmente, à tradução, as metilproteínas de arginina estão associadas à reação de oxidação-redução e metabolismo de carboidratos. Diferentemente, em *T. brucei* na sua forma procíclica, as proteínas metiladas em arginina estão principalmente associadas às funções do citoesqueleto e à locomoção.

Assim como na análise preliminar, percebe-se que os termos relacionados à localização celular (*Cellular Component*) das proteínas metiladas em *T. cruzi* estão escassamente anotados, pois 593 (67,5%) não possuem anotação. Ao serem analisados a classificação e enriquecimento funcional, a categoria mais representada, *integral component of membrane*, conta com 109 proteínas. Essa considerável identificação de proteínas de membrana deve-se, provavelmente, pela eficiência na etapa de extração de proteínas no protocolo FASP, através da utilização de SDS nesta etapa. Outra categoria bem representada possui 32 proteínas e são associadas aos ribossomos, categoria esta que está significativamente enriquecida dentro do metilproteoma identificado na forma epimastigota de *T. cruzi*. Associado a isso estão enriquecidos *Translation* como Processo Biológico e *Structural constituent of ribosome* como Função Molecular, demonstrando que a metilação exerce em *T. cruzi* um impacto direto na maquinaria de síntese proteica.

#### 5.4 O *crosstalk* entre metilação e outras modificações em *T. cruzi*

Desde que o código de combinações entre modificações pós-traducionais (PTMs) vem sendo estudado em histonas (STRAHL; ALLIS, 2000), o *crosstalk* entre as modificações vem sendo aos poucos desvendado, principalmente através da proteômica baseada em espectrometria de massas (OLSEN; MANN, 2013). Sabe-se, por exemplo, que a metilação em arginina afeta outros resíduos metilados, como as próprias lisinas, e outras modificações como acetilação e fosforilação (BEDFORD; CLARKE, 2009). Além disso, há uma inter-relação entre a metilação das lisinas 4 e 79 da histona H3 (H3K4 e H3K79) com a ubiquitinação da lisina 123 da histona H2B (H2BK123) em *Saccharomyces cerevisiae*, onde a não ubiquitinação de K123, previne a metilação de K4 e de K79 (FAROOQ et al., 2016). Há ainda, antagonismo entre metilação de arginina e a fosforilação de serina no domínio C-terminal da RNA polimerase II, o que impacta a transcrição de genes específicos em mamíferos (SIMS et al., 2011). Entre os efeitos da metilação em lisina e a interação com fosforilação, já foi reportado o *crosstalk* entre a metilação da lisina 810 (K810) na proteína supressora de tumor do retinoblastoma que, ao impedir o reconhecimento por uma quinase dependente de ciclina (Cdk) em resíduo de serina (S) adjacente, não permite a fosforilação deste substrato, controlando a progressão do ciclo celular (CARR et al., 2011).

Nossos dados revelam a coocorrência entre metilação e fosforilação em 62 proteínas de *T. cruzi*. Destas, 33 são proteínas hipotéticas conservadas e 2 são proteínas hipotéticas em *T. cruzi*. A análise de clusterização com alta estrigência realizada revela no cluster 1 proteínas envolvidas diretamente com o metabolismo de RNA (Figura 4.19). Sobressaem-se nesse cluster as *RNA-binding proteins* que são um dos tipos de proteínas mais reportadas como sendo metiladas em resíduos de arginina, apresentam os domínios de ligação a RNA (*RNA-binding domains*) também conhecidos como *RNA recognition motif* (RRM), já foram identificadas também em *T. brucei* (GOULAH; PELLETIER; READ, 2006; LOTT et al., 2013).

Apesar de ainda pouco estudada, a abrangência da metilação de proteínas nas células pode ser comparável com a ocorrência de modificações como a fosforilação e ubiquitinação, como demonstrado para células humanas recentemente (LARSEN et al., 2016). Da mesma forma, nossa análise do metilproteoma demonstra que a metilação parece ser tão abundante quanto a fosforilação em *T. cruzi*.

## 6 CONCLUSÕES

Com base nos resultados aqui expostos, conclui-se que:

- As análises *in silico* da maquinaria de metilação e demetilação em *T. cruzi* sugerem que: (i) *T. cruzi* possui metiltransferases de lisina (TcPKMTs), metiltransferases de arginina (TcPRMTs) e demetilases JmjC; (ii) essa maquinaria de metilação e demetilação de *T. cruzi* é conservada em relação a *T. brucei* e humanos;
- A metilação de proteínas (Me, Me2 e Me3) é amplamente encontrada em *T. cruzi*;
- O fracionamento por gel gradiente SDS-PAGE seguido de LC-MS/MS foi a estratégia mais eficiente na identificação de sítios metilados em arginina e lisina em *T. cruzi*;
- As proteínas metiladas de *T. cruzi* estão localizadas em diferentes compartimentos celulares e apresentam ampla gama de funções, sugerindo uma influência da metilação em diversos processos biológicos do parasito;
- As proteínas de *T. cruzi* metiladas em arginina estão envolvidas majoritariamente em processos de oxiredução e metabolismo de carboidratos;
- As proteínas de *T. cruzi* metiladas em lisina estão envolvidas majoritariamente na síntese de proteínas;
- A metilação em lisina e arginina apresenta coocorrência com fosforilação em 62 proteínas de *T. cruzi*;
- A metilação de proteínas parece ser tão abundante e importante quanto a fosforilação em *T. cruzi*.

## REFERÊNCIAS

- AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. **Nature**, v. 422, n. March, 2003.
- AEBERSOLD, R.; MANN, M. Mass-spectrometric exploration of proteome structure and function. **Nature**, v. 537, n. 7620, p. 347–355, 2016.
- AFJEHI-SADAT, L.; GARCIA, B. A. Comprehending Dynamic Protein Methylation with Mass Spectrometry. **Current opinion in chemical biology**, v. 17, n. 1, p. 12–19, 2014.
- AMATO NETO, V.; SHIKANAI-YASUDA, M. A.; AMATO, V. S. Doença de Chagas aguda. In: Dias, J. C. P. & Coura, J. R. **Clínica e terapêutica da doença e Chagas**, Editora Fiocruz, 1997, p.127-133.
- AMBLER, R. P.; REES, M. W. Epsilon-N-Methyl-lysine in Bacterial Flagellar Protein. **Nature**, v. 183, p. 1654–1655, 1959.
- ASHBURNER, M. et al. Gene Ontology : tool for the unification of biology. **Nature genetics**, v. 25, n. may, p. 25–29, 2000.
- BEDFORD, M. T.; CLARKE, S. G. Protein arginine methylation in mammals: who, what, and why. **Molecular cell**, v. 33, n. 1, p. 1–13, 16 jan. 2009.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289–300, 1995.
- BERRIMAN, M. et al. The genome of the African trypanosome *Trypanosoma brucei*. **Science**, v. 309, n. 5733, p. 416–22, 15 jul. 2005.
- BIGGAR, K. K.; LI, S. S.-C. Non-histone protein methylation as a regulator of cellular signalling and function. **Nature Reviews Molecular Cell Biology**, v. 16, n. 1, p. 5–17, 2014.
- BÖTTGER, A. et al. The oxygenase Jmjd6 - a case study in conflicting assignments. **The Biochemical journal**, v. 468, n. 2, p. 191–202, 2015.
- BREMANG, M. et al. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. **Molecular bioSystems**, v. 9, p. 2231–47, 2013.
- CAO, X.; ARNAUDO, A. M.; GARCIA, B. A. Large-scale global identification of protein lysine methylation in vivo. n. May, p. 477–485, 2013.
- CARLSON, S. M.; GOZANI, O. Emerging Technologies to Map the Protein Methylome. **Journal of molecular biology**, 5 maio 2014.
- CARR, S. M. et al. Interplay between lysine methylation and Cdk phosphorylation in growth control by the retinoblastoma protein. **The EMBO journal**, v. 30, n. 2, p. 317–327, 2011.
- CDC (Centers for Disease Control and Prevention). **Parasites - American Trypanosomiasis (also known as Chagas Disease) - Biology**. 2015. Disponível em: <<http://www.cdc.gov/parasites/chagas/biology.html>> Acesso em 23 maio 2016.

- 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.
- CHANG, B. et al. JMJD6 is a Histone Arginine Demethylase. **Science**, v. 318, p. 444–447, 2007.
- CHEN, G. et al. In vitro methylation by methanol: Proteomic screening and prevalence investigation. **Analytica Chimica Acta**, v. 661, n. 1, p. 67–75, 2010.
- CHENG, K. Y. et al. Immunoblot Assay Using Recombinant Antigens as a Supplemental Test To Confirm the Presence of Antibodies to *Trypanosoma cruzi*. **Clinical and Vaccine Immunology**, v. 14, n. 4, p. 355–361, abr. 2007.
- CLOOS, P. A. C. et al. Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. **Genes and Development**, v. 22, n. 9, p. 1115–1140, 2008.
- CONTRERAS, V. T.; MOREL, C. M.; GOLDENBERG, S. Stage specific gene expression precedes morphological changes during *Trypanosoma cruzi* metacyclogenesis. **Molecular and Biochemical Parasitology**, v. 14, n. 1, p. 83–96, 1985.
- COPELAND, R. A. Molecular pathways: Protein methyltransferases in cancer. **Clinical Cancer Research**, v. 19, n. 23, p. 6344–6350, 2013.
- COURA, J. R. The main sceneries of chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 277–282, 2015.
- COX, J.; MANN, M. Is Proteomics the New Genomics? **Cell**, v. 130, n. 3, p. 395–398, 2007.
- COX, J.; MANN, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. **Nature biotechnology**, v. 26, n. 12, p. 1367–1372, 2008.
- DAI, C.; GU, W. p53 post-translational modification: deregulated in tumorigenesis. **Trends in Molecular Medicine**, v. 16, n. 11, p. 528–536, 2010.
- DE GODOY, L. M. F. et al. Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. **Proteomics**, p. 2694–2703, 2012.
- DE JESUS, T. C. L. et al. Chromatin proteomics reveals variable histone modifications during the life cycle of *Trypanosoma cruzi*. **Journal of Proteome Research**, 2016.
- DIAS, J. C. P. Ecological Aspects of The Vectorial Control of Chagas ' Disease in Brazil Aspectos Ecológicos do Controle de Vetores para Doença de Chagas no Brasil. **Caderno de Saúde Pública**, v. 10, p. 352–358, 1994.
- EL-SAYED, N. M. et al. Comparative Genomics of Trypanosomatid Parasitic Protozoa. **Science**, v. 309, n. July, p. 404–409, 2005a.
- EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, n. 5733, p. 409–15, 15 jul. 2005b.
- FAROOQ, Z. et al. The many faces of histone H3K79 methylation. **Mutation research. Reviews in mutation research**, v. 6, n. 2, p. 356–372, 2016.

- FIGUEIREDO, L. M.; CROSS, G. A M.; JANZEN, C. J. Epigenetic regulation in African trypanosomes: a new kid on the block. **Nature reviews Microbiology**, v. 7, n. 7, p. 504–513, 2009.
- FISK, J. C. et al. A Type III Protein Arginine Methyltransferase from the Protozoan Parasite *Trypanosoma brucei*. **The Journal of biological chemistry**, v. 284, n. 17, p. 11590–11600, 2009.
- FISK, J. C. et al. Proteomic Analysis Reveals Diverse Classes of Arginine Methylproteins in Mitochondria of trypanosomes. **Molecular & cellular proteomics**, p. 302–311, 2013.
- FISK, J. C.; READ, L. K. Protein Arginine Methylation in Parasitic Protozoa. **Eukaryotic Cell**, v. 10, n. 8, 2011.
- GOULAH, C. C.; PELLETIER, M.; READ, L. K. Arginine methylation regulates mitochondrial gene expression in *Trypanosoma brucei* through multiple effector proteins. p. 1545–1555, 2006.
- GRISARD, E. C. et al. Characterization of *Trypanosoma rangeli* Strains Isolated in Central and South America : an Overview. v. 94, n. 2, p. 203–209, 1999.
- GSTAIGER, M.; AEBERSOLD, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. **Nature Reviews Genetics**, v. 10, n. 9, p. 617–627, 2009.
- GUO, A. et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. **Molecular & Cellular Proteomics**, v. 13, n. 1, p. 372–387, 2014.
- HAN, X.; ASLANIAN, A.; YATES, J. R. Mass spectrometry for proteomics. **Current opinion in chemical biology**, v. 12, n. 5, p. 483–90, out. 2008.
- HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, v. 37, n. 1, p. 1–13, 2009a.
- HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocols**, v. 4, n. 1, p. 44–57, 2009b.
- IVENS, A. C. et al. The Genome of the Kinetoplastid Parasite, *Leishmania major*. **Science**, v. 309, n. 5733, p. 436–442, 2005.
- JENSEN, O. Interpreting the protein language using proteomics. **Nature reviews. Molecular cell biology**, v. 7, n. 6, p. 391–403, jun. 2006
- KIM, S.; PAIK, W. K. Studies on the Origin of  $\epsilon$ -N-Methyl-L-lysine in Protein Studies on the Origin in Protein. **J. Biol. Chem.**, v. 240, n. 12, p. 4629–4634, 1965.
- KLOSE, R. J.; KALLIN, E. M.; ZHANG, Y. JmjC-domain-containing proteins and histone demethylation. **Nature reviews. Genetics**, v. 7, n. 9, p. 715–27, 2006.
- LANOUILLE, S. et al. The functional diversity of protein lysine methylation. **Molecular Systems Biology**, v. 10, p. 1–26, 2014.
- LARSEN, S. C. et al. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. **Science Signaling**, v. 9, n. 443, p. 1–15, 2016.



- LAWRENCE, A.-M.; BESIR, H. U. S. Staining of proteins in gels with Coomassie G-250 without organic solvent and acetic acid. **Journal of visualized experiments**, n. 30, p. 2–4, 2009.
- LOTT, K. et al. Global proteomic analysis in trypanosomes reveals unique proteins and conserved cellular processes impacted by arginine methylation. **Journal of proteomics**, v. 91, p. 210–25, 8 out. 2013.
- LUCENA, A. C. R. **Proteoma e fosfoproteoma quantitativo durante estresse nutricional da forma epimastigota de *Trypanosoma cruzi***. 2015 (Dissertação de mestrado).
- MARCHINI, F. K. et al. Profiling the *Trypanosoma cruzi* Phosphoproteome. **Plos One**, v. 6, n. 9, 2011.
- MARCHLER-BAUER, A. et al. CDD : NCBI ' s conserved domain database. **Nucleic acid research**, v. 43, n. November 2014, p. 222–226, 2015.
- MARTINEZ, R. J. On the chemistry of flagella of *Spirillum serpens*. **Biochemical and biophysical research communications**, v. 12, n. 3, 1963.
- MOORE, K. E. et al. A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. **Molecular Cell**, v. 50, n. 3, p. 444–456, 2013.
- MORALES, Y. et al. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). **Archives of Biochemistry and Biophysics**, v. 590, p. 138–152, 2016.
- MURRAY, K. The Occurrence of E-N-Methyl Lysine in Histones. **Biochemistry**, v. 3, p. 10–15, 1964.
- NEVES, D. P. **Parasitologia dinâmica**. São Paulo: Atheneu, 2003, p. 113-115.
- NIELSEN, M. L.; SAVITSKI, M. M.; ZUBAREV, R. A. Extent of Modifications in Human Proteome Samples and Their Effect on Dynamic Range of Analysis in Shotgun Proteomics. **Molecular & Cellular Proteomics**, v. 5, p. 2384–2391, 2006.
- OLSEN, J. V. et al. Higher-energy C-trap dissociation for peptide modification analysis. **Nature Methods**, v. 4, n. 9, p. 709–712, 2007.
- OLSEN, J. V; MANN, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. **Molecular & cellular proteomics : MCP**, v. 12, n. 12, p. 3444–52, 2013.
- ONG, S.-E.; MITTLER, G.; MANN, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. **Nature methods**, v. 1, n. 2, p. 119–26, nov. 2004.
- PAIK, W. K.; KIM, S. Enzymatic methylation of Protein fractions from calf thymus nuclei. **Biochemical and Biophysical Research Communications**, v. 29, n. 1, p. 14–20, 1967.
- PAIK, W. K.; PAIK, D. C.; KIM, S. Historical review: the field of protein methylation. **Trends in Biochemical Sciences**, v. 32, n. 3, p. 146–152, 2007.
- PATTERSON, S. D.; AEBERSOLD, R. H. Proteomics: the first decade and beyond. **Nature Genetics**, v. 33, n. 3s, p. 311–323, 2003.
- PEREZ, C. J.; LYMBERG, A. J.; THOMPSON, R. C. A. Reactivation of Chagas Disease: Implications for Global Health. **Trends in Parasitology**, v. 31, n. 11, p. 595–603, 2015.

RAPPSILBER, J.; ISHIHAMA, Y.; MANN, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption / Ionization , Nanoelectrospray , and LC / MS Sample Pretreatment in Proteomics. **Analytical Chemistry**, v. 75, n. 3, p. 663–670, 2003.

RAPPSILBER, J.; MANN, M.; ISHIHAMA, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. **Nature protocols**, v. 2, n. 8, p. 1896–1906, 2007.

ROSENZWEIG, D. et al. Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. **Proteomics**, p. 1843–1850, 2008.

SCHROEDER, M. J. et al. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. **Analytical Chemistry**, v. 76, n. 13, p. 3590–3598, 2004.

SIMS, R. J. et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. **Science (New York, N.Y.)**, v. 332, n. 6025, p. 99–103, 2011.

SIQUEIRA, A. M. Subfilo Mastigophora. In: NEVES, David P (org.). **Parasitologia Humana**. 11. ed. Porto Alegre: Atheneu, 2004. p. 37-40.

STEINDEL, M. et al. Characterization of *Trypanosoma cruzi* isolated from humans, vectors, and animal reservoirs following an outbreak of acute human Chagas disease in Santa Catarina State, Brazil. **Diagnostic microbiology and infectious disease**, v. 60, n. 1, p. 25–32, jan. 2008.

STENFLO, J. et al. Vitamin K dependent modifications of glutamic acid residues in prothrombin. **Proceedings of the National Academy of Sciences of the United States of America**, v. 71, n. 7, p. 2730–3, 1974.

STEVENS, J. R. Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. **Parasite**, v. 15, n. August, p. 226–232, 2008.

STRAHL, B. D.; ALLIS, C. D. The language of covalent histone modifications. **Nature**, v. 403, n. 6765, p. 41–45, 2000.

SYLVESTERSEN, K. B. et al. Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. **Molecular & cellular proteomics : MCP**, v. 13, n. 8, p. 2072–2088, 2014.

SYLVESTERSEN, K. B.; NIELSEN, M. L. Large-Scale Identification of the Arginine Methylome by Mass Spectrometry. **Current Protocols in Protein Science**, n. November, p. 24.7.1-24.7.17, 2015.

THERMO FISHER SCIENTIFIC. **Overview of Post-Translational Modifications (PTMs)**. 2016. Disponível em: < <https://www.thermofisher.com/br/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html>>.

TSCHIRSCH, B. et al. The protein encoded by the *Drosophila* position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. **The EMBO journal**, v. 13, n. 16, p. 3822–3831, 1994.

UHLMANN, T. et al. A Method for Large-scale Identification of Protein Arginine Methylation. **Molecular & cellular proteomics**, p. 1489–1499, 2012.

- VICKERMAN, K. The evolutionary expansion of the trypanosomatid flagellates. **International Journal for Parasitology**, v. 24, n. 8, p. 1179–1201, 1994.
- WALPORT, L. J. et al. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. **Nature Communications**, v. 7, n. May, p. 1–12, 2016.
- WANG, K. et al. Proteomic analysis of protein methylation in the yeast *Saccharomyces cerevisiae*. **Journal of Proteomics**, v. 114, p. 1–8, 2014.
- WASINGER, V. C. et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. **Electrophoresis**, v. 16, n. 7, p. 1090–1094, 1995.
- WHO (World Health Organization). **Chagas disease (American trypanosomiasis)**. World Health Organization, 2016. Disponível em: <<http://www.who.int/chagas>>.
- WIŚNIEWSKI, J. R. et al. Universal sample preparation method for proteome analysis (FASP). **Nature methods**, v. 6, n. 5, p. 359–362, 2009.
- WIŚNIEWSKI, J. R.; ZOUGMAN, A.; MANN, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. **Journal of Proteome Research**, v. 8, n. 12, p. 5674–5678, 2009.
- YANG, F. et al. High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. **Expert Review of Proteomics**, v. 9, n. 2, p. 129–134, 2012.
- ZHANG, X. et al. Emerging roles of lysine methylation on non-histone proteins. **Cellular and Molecular Life Sciences**, v. 72, n. 22, p. 4257–72, 2015.
- ZHANG, X.; WEN, H.; SHI, X. Lysine methylation: Beyond histones. **Acta Biochimica et Biophysica Sinica**, v. 44, n. 1, p. 14–27, 2012.

## ANEXOS

### Anexo 1- Alinhamentos das PRMTs, PKMTs e demetilases em *Trypanosoma cruzi* CLBrener Esmeraldo e Non Esmeraldo Like

#### PRMTs

<p>TcPRMT1_EL_TcCLB.506529.50 TcPRMT1_NEL_TcCLB.508593.110</p>	<pre> 1 MKNRRKKKKEKRRTENAHHKSSRALPFVSVCFVFFLLKWDWEGTFVFNFFSFLSFFFVFPVPPPSVADISLLRRIQRASFPFATARKTTSQTALLASQKE 1 .....MTTITETPAPTK 112 1 .....MTTITETPAPTK 111                 </pre>	
<p>TcPRMT1_EL_TcCLB.506529.50 TcPRMT1_NEL_TcCLB.508593.110</p>	<pre> 113 APVAGSATGDYYFDSYSHYG IHMEMLKDYHRTTSYRDAIWRNAYLFKGVVLDVGCCTGILSMFAAKAGARKVIGVDCSSVTVQARQIVKDNFEGVITIIQGKMEDLQLE 12 APVAGSATGDYYFDSYSHYG IHMEMLKDYHRTTSYRDAIWRNAYLFKGVVLDVGCCTGILSMFAAKAGARKVIGVDCSSVTVQARQIVKDNFEDVITIIQGKMEDLQLE 224 123                 </pre>	<p>Identidade: 75,7% Posições idênticas: 334</p>
<p>TcPRMT1_EL_TcCLB.506529.50 TcPRMT1_NEL_TcCLB.508593.110</p>	<pre> 225 KVDIII SEWMGYFLLYESMLNTVLYARDRWGTPDVKLPDRANMYACGISDQQYKEERFDVWDDVQGFDFSYFKRLSYIEPLIDTVDRNQILTNMSNLSFNINTVQEAELA 124 KVDIII SEWMGYFLLYESMLNTVLYARDRWGTPDVKLPDRANMYACGISDQQYKEERFDVWDDVQGFDFSYFKRLSYIEPLIDTVDRNQILTNMSNLSFNINTVQEAELA 336 235                 </pre>	
<p>TcPRMT1_EL_TcCLB.506529.50 TcPRMT1_NEL_TcCLB.508593.110</p>	<pre> 337 FTSEFELIAERSGCVDIAISVHFDTFPAAGHEVIVLTTTLEAPTHWRQTVLYLFINPLHMNKGEKAKFRMSCCPNKGPNRDLIALHVDFDGGKQSSHYDQDFRLR 238 FTTEFELIAERSGCVDIAISVHFDTFPAAGHEVVVLTPTLEAPTHWRQTVLYLFINSHMKNKGEKAKFRMSCCPNKGPNRDLIALHVDFDGGKQSSHYDQDFRLR 441 340                 </pre>	
<p>TcPRMT3_EL_TcCLB.510311.140 TcPRMT3_NEL_TcCLB.508153.1110</p>	<pre> 1 MPSRKNPPPYELGQGQQRKKKVFNTKNVANGACSPTRTVANESLLATEGEKNATSKSLYGDPAAARISNLYSISDRIRLRAYQSVLKGIKGKSLHIGCGMGLISLMAAR 1 MPSRKNLPPPYELGQGQQRKKKVFNTKNVTNGACSPTRTVANESLLATEGEKNATSKSLYGDPAAARISNLYSISDRIRLRAYQSVLKGIKGKSLHIGCGMGLISLMAAR 111 111                 </pre>	
<p>TcPRMT3_EL_TcCLB.510311.140 TcPRMT3_NEL_TcCLB.508153.1110</p>	<pre> 112 NLASAVVAIDHSAIVDAAQVVAQKQGLQNSFFRGLREMMGNFPVKKFDVILCEWMGTFVNDPLEDALFAREHLLAENGVMCPDSSIHIVGVSDYAFYLDTVEYWSN 112 NLASAVVAIDHSAIVDAAQVVAQKQGLQNSFFRGLREMMGNFPVKKFDVILCEWMGTFVNDPLEDALFAREHLLAENGVMCPDSSIHIVGVSDYAFYLDTVEYWSN 222 222                 </pre>	<p>Identidade: 98,44% Posições idênticas: 379</p>
<p>TcPRMT3_EL_TcCLB.510311.140 TcPRMT3_NEL_TcCLB.508153.1110</p>	<pre> 223 VYGFKMDPMKGLVRQEVETCAIPASNIVTSTSLIHTVNI TEFRDADNKDRDESGETVPFVSRATCNTTVNYLTFVVDATFTNPHDPGANFVIGIRPGGNNAWTETSVGLPQ 223 VYGFKMEPMKGLVRQEVETCAIPASNIVTSTSLIHTVNI TEFRDADNKDADESGETVPFVSRATRNTTVNYLTFVVDATFTNPHDPGANFVIGIRPGGNNAWTETSVGLPQ 333 333                 </pre>	
<p>TcPRMT3_EL_TcCLB.510311.140 TcPRMT3_NEL_TcCLB.508153.1110</p>	<pre> 334 PLPLKCGETLSGELRVRVLRNRGQVTLVEVTAQTDGKVLAEKTTGSYTYQRF 334 PLPLKCGETLSGELRVRVLRNRERGVTLVEVTAQTDGKVLAEKTTGSYTYQRF 385 385                 </pre>	

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

1 MESVAATAASMANERKRRGFESRTAWLMPSEEGVFQSLSLRHFGFTAGKMLQFALQQQAACGASMCVFPIDEAAADGSSGRQTTGCSPLAALASNLDMKSGMDQTWGLLSR 113

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

114 VGSWGFSVCGRVTASSWDLADERSPFGLRLRAAFLGLSAVLVPLPNDVTTPAFTDAVTVLTDHLRQNRITNIWVYCDASSQAHRVAFSRLRTALMWGDAPPSLTDAAGCV 226

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

227 AREALHRVTPYLYFSETFGALPSEWLGEVVAFEVPHIDKLMRALRSGPFHLEPDSNLPKIDLPLINSYWATAQPPFISLAAFIVELLRRAAIFSVKDFIPAYTLLNQLYE 339

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

1 .....MVRADHDERGHGRLCSKYLKNTGGDKMDVAVVLLGAGRGLIAECLAASSIGV 56  
340 RHVLDNRDNFSSFEDIQLPLQLPLGHMLSSGVYEVFECDFVKYQRYHAAMLNYFREWLEHDERGHGRLCSKYLKNTGGDKMDVAVVLLGAGRGLIAECLAASSIGIRV 452

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

57 HLFAVEKNPEAMEFICLRLRSDPQWRNYISACGHLVETMCADGRAVTATIEGPLPKLWGLCDLVVSELLGSGFDNELSPECIDGFYEDFRRYQNSMGIPPNHMHVSIPOEYTA 169  
453 HLFAVEKNPEAMEFIRLRLRSDPQWYNIISACGHLVETMCADGRAVTATIEGPLPKLWGLCDLVVSELLGSGFDNELSPECIDGFCEDFRRYQNSMGIPPNHMHVSIPOEYTA 565

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

170 WIAPLYSGRMEESLADTAVSGLTVLLPWCHDRRAAIFHSMFVTDLCRQVSLSTPQPQWTFQHFTRDFQKEREALLTRMNTAGRCSGFIGYFTAVLFASKSNGADNVGNVD 282  
566 WIAPLYSGRMEESLADTAVSGLTVLLPWCHDRRAAIFHSMFVTDLCRQVSLSTPQPQWTFQHFTRDFQKEREALLTRMNTAGRCSGFIGYFTAVLFASKSNGADNVGNVD 678

TcPRMT6\_EL\_TcCLB.508781.129  
TcPRMT6\_NEL\_TcCLB.506853.30

283 ISTTLSTTPEERTPEMFSWFPCCFFAVEPRDVLEVEADDRLQLRLRRRCVND EKRVFYSYDARLLKGHEGEVKATTAVINDG6WAASILLTS 373  
679 ISTTLSTTPEERTPEMFSWFPCCFFAVEPRDVLEVEADDRLQLHLRRRCVNEEKRVFYSYDARLLKGREGEVKATTAVINDG6WAASILLNS 769

TcPRMT6\_EL\_TcCLB.506947.80  
TcPRMT6\_NEL\_TcCLB.507057.30

1 MFLWKRFMPLQQFAEDYNPHESAQEQQYDAYADLGVHRVMSDAPRMSFYRQAVTCTETVVQDRVVVDVGS6TGIFSIWAARAGAKHVFSIEASSLSFFQARVIDDNHLSDKIT 113  
1 MFLWKRFMPLQQFAEDYNPHESAQEQQYDAYADLGVHRVMSDAPRMSFYRQAVTCTETVVQDRVVVDVGS6TGIFSIWAARAGAKHVFSIEASSLSFFQARVIDDNHLSDKIT 113

TcPRMT6\_EL\_TcCLB.506947.80  
TcPRMT6\_NEL\_TcCLB.507057.30

114 LLSETVENIVAGGVASFIATHEAVLKGC6VAVIVSEWMGFYLLHEGMLPSVIRARDFGGEVNAKLGVCRRRIEMIPERATIRVAPMTCAPYCEERYCRFWSGLEGVDFSRYGTV 226  
114 LLSETVENIVAGGVVSFIATHEAVLKGC6VAVIVSEWMGFYLLHEGMLPSVIRARNFGEVNAKLGVCRRRIEMIPERATIRVAPMTCAPYCEERYCRFWSGLEGVDFSRYGTV 226

TcPRMT6\_EL\_TcCLB.506947.80  
TcPRMT6\_NEL\_TcCLB.507057.30

227 ECEMQLETSNPLIDTIPASCLLHEGVVLAEVELSSIAEDAMTSLCRTVHFNFONSTAFQRHFVNDCTQEVAVDGFLLWFDVSFRWTLNTPSPFHPPTHWKQTTVLLPATVRAE 339  
227 ECEMQLETSNPLIDTIPASCLLHEGVVLAEVELSSIAEDAMTSLCRTVHFSLONSTAFQRHFVNDCTQEVAVDGFLLWFDVSFRWTLNTPSPFHPPTHWKQTTVLLPATVRAE 339

TcPRMT6\_EL\_TcCLB.506947.80  
TcPRMT6\_NEL\_TcCLB.507057.30

340 RLVAFTNSDATLDAEVRVAVATGLRRRCYTIIEFELK 374  
340 RLVAFTNSDATLDAEVRVAVATGLRRRCYTIIEFELK 374

Identidade: 46,3%  
Posições idênticas: 356

Identidade: 97,6%  
Posições idênticas: 365

TcPRMT7\_EL\_TcCLB.509153.100 1 MPARWHNKQKKGSNGHKRPS TGGNGSAARRDIGFIPPGASAVQNSSAHADTFMYRFQNFVPRRLRHAPTSLVEAVNDFHFAMMND RPRNEFYELLKRHVTPETGVLEI 112  
TcPRMT7\_NEL\_TcCLB.510303.310 1 MPARWHNKQKKGSNGHKRPTGGNGSAARRDIGFIPPGASAVQNSSAHADTFMYRFQNFVPRRLRHAPTSLVEAVNDFHFAMMNDQPRNEFYELLKRHVTPETGVLEI 112

TcPRMT7\_EL\_TcCLB.509153.100 113 GAGSGLLSLMAAQLGAKWVVAVEGSAEMAKLARANICANKMEHKVTVLHMLSTELTLRHLPDRPDVMVSEILGTLLEGESALDYIADVRELLKPTTKIIPQHGTQYAVPIE 224  
TcPRMT7\_NEL\_TcCLB.510303.310 113 GAGSGLLSLMAAQLGAKWVVAVEGSAEMANLARANICANKMEHKVTVLHMLSTELTLRHLPDRPDVMVSEILGTLLEGESALDYIADVRELLKPTTKIIPQHGTQYAVPIE 224

Identidade: 97,4%  
Posições idênticas: 408

TcPRMT7\_EL\_TcCLB.509153.100 225 CDTLRDVCVAVS6WKGLDLSHMTALQDQTVSIVFAKQYGFMRSSVFPFRLSEPIALFVDFASAVRRDIPMNLKLDVKATSSGTAHAMLYYWKAFDGLVMSTDPADTLNNFPR 336  
TcPRMT7\_NEL\_TcCLB.510303.310 225 CDTLRDVCVAVS6WKGLDLSHMTALQDQTVSIVFAKQYGFMRSSVFPFRLSEPIALFVDFASAVRRDIPMNLKLDVKATSSGTAHAMLYYWKAFDGLVMSTDPADTLNNFPR 336

TcPRMT7\_EL\_TcCLB.509153.100 337 DMQWQALQLLDAANGPFPPTPVVFTGEELKFQCHFSSDRVTVAHILCPTAEAAASPAKTTPTDNDAIKTEKEACEKVNK6KK 419  
TcPRMT7\_NEL\_TcCLB.510303.310 337 DMQWQALQLLDAANGPFPPTPVVFTGEELKFQCHFSSDRVTVAHILCPTAEAAASPAKTTPTDNDAIKTENEACEKVNK6KK 419

## PKMTs

DOT1\_EL\_TcCLB.510431.70 1 MTVQRYLLGGALFFFVYCLVRPGLRRLAFASGIGFALGYGLPALNNVRKHSFKAINEPHLLAPPADFRSMYHRIFEHISERYDELGQMEIGEEHITVMQNVFAQIGTLDFD 115  
DOT1\_NEL\_TcCLB.460125.10 1 MTVQRYLLGGALFFFVYCLVRPGLRRLAFASGIGFALGYGLPALNNVRKHSFKAINEPHLLAPPADFRSMYHRIFEHISERYDELGQMEIGEEHITVMQNVFAQIGTLDFD 115

DOT1\_EL\_TcCLB.510431.70 116 GIRKMWTSVILPQLRSSISGGALVNDSQLPTKIEESVVFCDIGSGVGNVCLQILAETHCKKVVGVVEIPSRFRSAKIALANAKFYYPEYFDGKDAIFFNDDFVNCVSLKEQQV 230  
DOT1\_NEL\_TcCLB.460125.10 116 GIRKMWTSVILPQLRSSISGGALVNDSQLPTKIEESVVFCDIGSGVGNVCLQILAETHCKKVVGVVEIPSRFRNAKTALANAKLYYPEYFDGKEEIFFNDDF..... 218

Identidade: 70,43%  
Posições idênticas: 212

DOT1\_EL\_TcCLB.510431.70 231 NVIFAHSWMFDELMAKLTELVTSVPTVLGVVTSRRLTEKVLQISPIKLVSQVYLTADWNEEAPFYVYTKK 301  
DOT1\_NEL\_TcCLB.460125.10 ..... 301

# Demetilases

*JmjD4\_EL\_TcCLB.506369.30* 1 .....MLELDGRTLTFEKFRECCLEPNVPAVIRCAVPECETGSKGQFVSCLERMTSRLAPMSLIDAYGEGHHVPTVEFMTGKTSDSDQFECMECRM 92  
*JmjD4\_NEL\_TcCLB.510225.50* 1 MGFVGVGERGKMCVEVKQGGGTCVLELDGRTLTFEKFRECCLEPNVPAVIRCAVPECETGSKGQFVSCLERMTSRLAPMSLIDAYGEGHHVPTVEFMTGKTSDSDQFECMECRM 115

*JmjD4\_EL\_TcCLB.506369.30* 93 PLGEVMECWRRKDGVLKLDWHMQLSESQGGATERN SVEKKAVIDGYHCIVHGDLQVPEYLGHDMWDDFCRCLGGGEANYCHFGDTESDYRFAYIGPPSSWTLPHDFVFGTY 207  
*JmjD4\_NEL\_TcCLB.510225.50* 116 PLGEVMECWRRKDGVLKLDWHMQLSESQGGATERTSVEKKAMIDGYHCIVHGDLQVPEYLGHDMWDDFCRCLGGGEANYCHFGDTESDYRFAYIGPPSSWTLPHDFVFGTY 230

*JmjD4\_EL\_TcCLB.506369.30* 208 SWSLNVCGEKLWFFPSPEGNQRLLLETSLYGLMLPVDIRTTTGAELWAVLQRPGLVFPVPSGYLHQVHNVSGPQFLPPKNTLRPDDANPKKAKVLDNDDTGSVSLVISIHNHWG 322  
*JmjD4\_NEL\_TcCLB.510225.50* 231 SWSLNVCGEKLWFFPSPEGNQRLLLETSLYGLMLPVDIRTTTGTTELWAVLQRPGLVFPVPSGYLHQVHNVSGPRFLPPKNTLRPDDANPKKAKVLDNDDTGSVSLVISIHNHWG 345

*JmjD4\_EL\_TcCLB.506369.30* 323 NEWCVERMVEAFCRDANRLQRLLEEDRVLLFGEDCTAWHKHVENLLIGGTNNWFACIRSFCLKHRLACLSDGKGAEVGLRRLKCLTEVDAVELRITHTCGS 428  
*JmjD4\_NEL\_TcCLB.510225.50* 346 NEWCVERMVEAFCRDANRLQRLLEEDRVLLFGEDCTAWHKHVENLLIGGTNNWFACIRSFCLKHRLACLSDGKGAEVGLRRLKCLAEVDAVELRITHTCGS 451

*JmjD7\_EL\_TcCLB.506289.60* 1 MTDLCGVRYLCFCFSFLSFFFFFCVVCVCCVCIISTRCSHTWPRHDFPLRIVCLFVVCCWFLFFISVCMQGAAGREEGRVGGRTMTDVVEAPRPATVAEAFANTAKALEQFNIE 115  
*JmjD7\_NEL\_TcCLB.508881.30* 1 .....MTDVVEAPRPATVAEAFANTAKALEQFNIE 30

*JmjD7\_EL\_TcCLB.506289.60* 116 RIRVEGLS...STHCWKHKNDMEKEEDVRTLTEQAFMREFVFQSRPCVIVGALEDWPAMQRWRDDRYLFDLDGHLGVRRETDDAAAPSSETAATGDVAVVPKNDHSSDD 226  
*JmjD7\_NEL\_TcCLB.508881.30* 31 RIRVEGLSCKSSI THCRKHDKNNMEKEEDVRTLTEQAFMREFVFQSRPCVIVGALEDWPAMQRWRDDRYLFDLDGHLGVRRETDDAAAPSSETAATGDVAVVPKNDHSSDD 145

*JmjD7\_EL\_TcCLB.506289.60* 227 STHPREEEEEKDLKRVTVALTTPNGRADAVTYVYQGAREEGRENDADGKDNVEEHFKTEKLFMAAAEIRATLPELYHLLRENPLYPPVRPVFVDLRRPAPVVAYAQMNQCL 341  
*JmjD7\_NEL\_TcCLB.508881.30* 146 NTHPREEEEEKDLKRVTVALTTPNGRADAVTYVYQGAREEGRENDVDGKDNVEEHFKTEKLFMAAAEIRATLPELYHLLRENPLYPPVRPVFVDLRRPAPVVAYAQMNQCL 280

*JmjD7\_EL\_TcCLB.506289.60* 342 EVEYQHLHGDIRPELDRFGERVFGPEHEAANVWF6TPASVSSMHQDWWENLYAVVRGVKEFILLPPWEGVFPKPELPSAAFVQRQDGVNGYVFS6TPVHDGLTIPWMDVEFTP 458  
*JmjD7\_NEL\_TcCLB.508881.30* 261 EVEYQHLHGDIRPELDRFGERVFGPEHEAANVWF6TPASVSSMHQDWWENLYAVVRGVKEFILLPPWEGVFPKPELPSAAFVQRQDGVNGYVFS6TPVHDGLTIPWMDVEFTP 375

*JmjD7\_EL\_TcCLB.506289.60* 457 AALEDVSLMPTNMQERLHPLVVYVQPGVEVLYLPAMWHLRVAQHADDTRLARETQLTSSLQSTQDPPPLPIAAVNYWYDMSFTNPVAVMLREFGLLL 553  
*JmjD7\_NEL\_TcCLB.508881.30* 376 AALEDVSLMPTNMQERLHPLVVYVQPGVEVLYLPAMWHLRVAQHADDTRLAREKQLTSSLQSAQDPPPLPIAAVNYWYDMSFTNPVAVMLREFGLLL 472

**Identidade: 92,23%**  
 Posições idênticas: 416

**Identidade: 81,86%**  
 Posições idênticas: 456

