GILBERTO SANTOS DE OLIVEIRA

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

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

São Paulo
2018

GILBERTO SANTOS DE OLIVEIRA

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

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

Área de concentração: Biologia da Relação Patógeno-Hospedeiro

Orientador: Prof. Dr. Giuseppe Palmisano

VERSÃO CORRIGIDA

São Paulo
2018

RESUMO

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

A doença de Chagas é uma das doenças negligenciadas mais importantes com um número estimado de 12 milhões de indivíduos infectados, a maioria vivendo na América Central e do Sul. O parasito protozoário *Trypanosoma cruzi* (*T. cruzi*) é o agente etiológico da doença de Chagas. O *T. cruzi* é geneticamente diversificado. Uma nova nomenclatura foi adotada para classificar as cepas de *T. cruzi* (TcI-TcVI e Tcbat), denominados *Discrete Typing Units* (DTUs), com base em suas características bioquímicas, imunológicas e fenotípicas. As DTUs de *T. cruzi* foram correlacionadas a diversos desfechos clínicos, destacando a importância de testes epidemiológicos e moleculares específicos. Apesar do desenvolvimento de métodos de caracterização de *T. cruzi* baseados em assinaturas genéticas, cada método apresenta vantagens e desvantagens. O nosso trabalho mostra a aplicação da espectrometria de massas para o ensaio de caracterização de cepas de *Trypanosoma cruzi* usando bibliotecas espectrais MS² de peptídeos (*Trypanosoma cruzi* Strain Typing Assay using MS², Tc-STAMS²). A novidade do método é o uso de espectros de fragmentação de peptídeos como impressões digitais específicas para classificar e identificar as DTUs. Inicialmente, uma biblioteca de espectros é gerada a partir das DTUs de *T. cruzi*. A biblioteca é subsequentemente desafiada usando espectros MS/MS das DTUs desconhecidas e confiavelmente atribuí a uma DTU específica em uma abordagem automática e computacional. Utilizando o método Tc-STAMS² foram testadas diferentes variáveis como preparação e tipo de amostra, configuração do instrumento e plataforma de identificação, proporcionando alta confiança e robustez na caracterização das DTUs de *T. cruzi*. O método Tc-STAMS² representa uma estratégia complementar aos métodos atuais de genotipagem de *T. cruzi* baseados em DNA. Além disso, o método permite a identificação de características específicas das DTUs que poderiam estar relacionadas à biologia de *T. cruzi*.

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

ABSTRACT

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

Chagas disease is one of the most important neglected diseases with an estimated number of 12 million infected individuals, the majority living in Central and South America. The *Trypanosoma cruzi* (*T. cruzi*) protozoan parasite is the etiological agent of Chagas' disease. *T. cruzi* is highly genetically diverse and a new nomenclature separated each strain into seven different genetic groups (TcI-TcVI and Tcbat), named Discrete Typing Units (DTUs) based on their biochemical, immunological and phenotypical characteristics. *T. cruzi* DTUs have been correlated to diverse clinical outcomes highlighting the importance of molecular epidemiological screens. Despite the development of *T. cruzi* typing methods based on genetic signatures, each method presents its own advantages and challenges. The work presented here shows the application of mass spectrometry for *Trypanosoma cruzi* Strain Typing Assay using MS² peptide spectral libraries (Tc-STAMS²). The novelty of the method is based on the use of peptide fragmentation spectra as strain-specific fingerprints to classify and identify DTUs. Initially, a spectra library is generated from characterized *T. cruzi* strains. The library is subsequently inspected using MS/MS spectra from unknown strains and confidently assigned to a specific strain in an automated and computationally driven approach. The Tc-STAMS² method was employed to test several variables such as sample type and preparation, instrument setup and identification platform. Tc-STAMS² provided high confidence and robustness in *T. cruzi* strain typing. The Tc-STAMS² method represents a proof-of-concept of a complementary strategy to the current DNA-based *T. cruzi* genotyping methods. Moreover, the method allows the identification of strain-specific features that could be related to the biology of *T. cruzi* strains and their clinical outcomes.

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

1 INTRODUÇÃO

1.1 Trypanosoma cruzi

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

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

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

1.2 Prevalência

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

1.3 Mortalidade

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

coincide com o surto de transmissão oral ocorrido em Santa Catarina no ano de 2005 (15). A letalidade continuou elevada em 2006 (5,9%), porém diminuindo constantemente ao passar dos anos (15).

1.4 Gênero *Trypanosoma*

O *Trypanosoma cruzi* pertence ao filo Euglenozoa, do qual fazem parte a ordem dos Kinetoplastídeos Euglenoidea e Diplonemea (16). A característica da ordem dos Kinetoplastídeos é sua mitocôndria única, o kDNA (17). O kDNA contém em sua composição moléculas dupla-fita circulares denominadas maxicírculos e minicírculos que formam uma rede, sendo responsável pela codificação de proteínas ou RNAs mitocondriais que são responsáveis pela edição dos genes transcritos (18).

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

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

1.5 Ciclo de Vida do *Trypanosoma cruzi*

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

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

diferenciam em tripomastigotas sanguíneas e em seguida, são liberados na corrente sanguínea para infectar novas células e refazer o ciclo (**Figura 2**) (4).

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

1.6 Discrete Typing Units

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

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

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

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

Anteriormente à atribuição das DTUs, os isolados de *T. cruzi* eram classificados de acordo com alguns critérios, como por exemplo, sua morfologia. Outros critérios como virulência, suscetibilidade a fármacos, patogenicidade e perfil eletroforético de isoenzimas (39-43) também foram utilizados. Além disso, estudos utilizando o perfil eletroforético de isoenzimas permitiu a separação de três grupos principais, nomeados como Zimodemas (44,

45). Posteriormente, Tibayrenc e Ayala com a utilização de maior número de isolados foram capazes de aumentar significativamente o número de zimodemas, que agora conta com 43 zimodemas (43). Estas evidências reforçaram a ideia que o *T. cruzi* é um organismo diplóide, que sua reprodução ocorre majoritariamente por divisão binária (reprodução assexuada) e que a reprodução por processo de hibridização (reprodução sexuada) é rara. Esses fatos mostram a estabilidade do organismo que ocorreu ao longo do tempo (45, 46).

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

As metodologias mencionadas anteriormente demonstraram que algumas DTUs consistem de linhagens híbridas. Como revisado por Zingales et al (36), as linhagens III e IV podem ter se formado a partir de um ancestral em comum entre I e II. Adicionalmente, as linhagens V e VI podem ter ocorrido pela combinação entre II e III (52-55). Portanto, conclui-se que, embora raros, trocas genéticas entre as cepas podem acontecer (56).

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

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

Atualmente, os métodos utilizados para tipificação das linhagens de *T. cruzi* são baseados na caracterização dos genes do ribossomo, 24S α -rRNA e 18S rRNA. Outras análises envolvem amplificação do DNA genômico (59), porém, necessita padrões complexos eletroforéticos (60). Esses métodos são capazes de discriminar as linhagens, entretanto, as interpretações dos resultados podem levar ao erro, como por exemplo a interpretação das análises de diferenças entre bandas de géis de eletroforese (61). Outro método introduzido recentemente é a MLST (Tipificação de Sequências Multilocus) que tem a capacidade de discriminar genótipos únicos em DTUs, porém, a sua implementação não é

simples para uma caracterização rápida em grandes coleções de isolados (59). Métodos simples de caracterização como o RFLP (Polimorfismo de Comprimento de Fragmento) e o MTq-PCR (PCR em Tempo Real Multiplex) têm sido aplicados em amostras de isolados clínicos demonstrando seus potenciais em aplicações na clínica (62).

1.7 Genômica e expressão gênica do *T. cruzi*

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

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

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

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

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

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

Nos eucariotos a transcrição ocorre da seguinte maneira: a RNA polimerase II liga-se ao gene promotor iniciando a transcrição do gene (80). O processo acaba resultando no

transcrito primário ou pré-RNA, contendo sequências de íntrons e éxons. Com a adição de 7-metilguanosina (m7G cap) na extremidade 5' do RNA, ocorre a remoção dos íntrons por *cis-splicing* e adição da cauda de adenina (poli-A) na extremidade 3', gerando o mRNA maduro (73, 81). Por outro lado, a transcrição em tripanossomatídeos (**Figura 4**) ocorre pelo reconhecimento das regiões promotoras tanto pela RNA polimerase I quanto a RNA polimerase III (78, 82-84). Porém, para a maioria dos genes, não existe sequência consenso para a polimerase II. Sendo a RNA polimerase II de baixa especificidade e devido a ambas as fitas do DNA poderem ser transcritas (72), a RNA polimerase inicia a transcrição indiscriminadamente no genoma e forma longos RNAs policistrônicos sem íntrons (85). Com isso, esses pré-RNAs policistrônicos são processados por poliadenilação na extremidade 3' e *trans-splicing* na extremidade 5' pela adição de uma sequência líder de aproximadamente 39 nucleotídeos chamada de *spliced leader* (**Figura 4**) (86). Ao contrário do que ocorre nos procariotos, nos tripanossomatídeos as unidades monocistrônicas maduras codificam proteínas que não pertencem a uma mesma via metabólica e os níveis de mRNAs são modulados por eventos pós-transcricionais (73).

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

1.8 Doença de Chagas

1.8.1 Fase Aguda

Esta fase inicia-se após a infecção das formas metacíclicas em hospedeiros humanos. A infecção pode ocorrer pela mucosa dos olhos (sinal de Romaña) ou por ferimento causado pelo inseto ou pelo ato de coçar (chagoma de inoculação) (**Figura 5**). Caso a contaminação ocorra por via oral, congênita ou por transfusão sanguínea não haverá sinal aparente (88). A fase aguda é normalmente assintomática, porém manifestações não específicas podem ser observadas, como febre, náusea, vômito, anorexia, miocardite, linfadenopatia e hepatoesplenomegalia. Esses sintomas iniciam-se de 1 a 2 semanas após a picada do inseto infectado, ou após alguns meses se a transmissão ocorrer por transfusão de sangue infectado. Após esse período, os anticorpos IgM são encontrados em níveis elevados no soro dos pacientes portadores da DC (89) e também são detectados anticorpos IgG nessa fase (90). O diagnóstico é realizado em menos de 10% dos casos, devido aos sintomas não específicos, e mais de 95% dos casos curam-se espontaneamente (91).

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

1.8.2 Fase Crônica

Após a fase aguda, 60-70% dos pacientes entram na fase latente, na qual não há desenvolvimento dos sinais clínicos da doença. Essa fase pode durar anos ou a vida inteira

do paciente. Os outros 40-30% desenvolvem a fase crônica, que caracteriza-se por comprometimento cardíaco ou digestivo (megacólon e megaesôfago) como revisto em Rassi A Jr (4). Além disso, A DC pode comprometer o sistema nervoso central e periférico (93). Essa fase pode iniciar-se 10-30 anos após o contato com o parasito.

1.8.3 Tratamento

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

1.9 Diagnóstico da Doença de Chagas

O diagnóstico da doença de Chagas na fase aguda normalmente não ocorre, devido aos sintomas dessa fase serem inespecíficos, prejudicando o tratamento da doença (90). O teste para a fase aguda baseia-se em exame parasitológicos diretos, nos quais uma grande quantidade de tripomastigotas pode ser visualizada por microscopia utilizando sangue fresco com anticoagulantes (96). Para o diagnóstico da fase crônica, é realizado o teste sorológico para detectar anticorpos contra *T. cruzi*, por exemplo, testes de IFI (Teste de Imunofluorescência Indireta), HAI (Teste de Hemaglutinação Indireta) e o teste ELISA (Ensaio

de Imunoabsorção Enzimática). Esses testes apresentam alta sensibilidade, porém baixa especificidade por apresentar reação cruzada com *Leishmania sp.* e *Trypanosoma rangeli* (97). A PCR (Reação em Cadeia da Polimerase) não é utilizada para diagnóstico clínico e sim em pesquisa laboratorial, devido à falta de padronização dos métodos (4). O diagnóstico da doença de Chagas congênita é realizado utilizando-se o cordão umbilical ou sangue venoso do recém-nascido, sendo um método muito invasivo (98). Embora os métodos apresentados sejam utilizados rotineiramente no laboratório para o diagnóstico da doença de Chagas, nenhum deles apresenta sensibilidade e precisão de 100%, tornando necessário o desenvolvimento de novas metodologias para a caracterização da doença.

1.10 Proteômica Baseada em Espectrometria de Massas

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

O estudo do transcriptoma (conjuntos de transcritos) pode analisar o estado pontual da célula em determinado tempo, condição, modificações pós-transcricionais ou *splicing* alternativos, porém, essas análises não irão fornecer informações sobre as proteínas expressas pelo fato de ocorrerem inúmeras modificações pós-traducionais (fosforilação, glicosilação, ubiquitinação, arginilação, etc) (102, 103). A proteômica é utilizada como ferramenta para a caracterização detalhada do proteomas, suas modificações pós-traducionais (PTMs), quantificação de todas as proteínas de uma amostra como também a interação entre proteínas (104).

O surgimento da proteômica como a conhecemos hoje teve início na década de 1980, com a utilização da separação de proteínas por gel bidimensional (*two-dimensional gel electrophoresis, 2DE*) (105, 106). Essa técnica consiste na separação de proteínas utilizando

dois parâmetros, o ponto isoelétrico (carga) e a massa molecular, e ambos são definidos baseados na mobilidade das proteínas em uma matriz de gel de poliacrilamida (106). Mesmo sendo o 2DE uma metodologia poderosa para identificação de proteínas, a mesma tem limitações, como por exemplo, a dificuldade em identificar proteínas de membranas, estando limitado a detecção de proteínas associadas a membranas e proteínas com baixa hidrofobicidade (107).

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

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

permitindo quantificação e identificação de peptídeos com maior velocidade e precisão (116). Adicionalmente, com a rápida melhora nos equipamentos de cromatografia e espectrômetros de massas e o desenvolvimento de algoritmos mais precisos nas análises dos dados gerados, a proteômica baseada em espectrometria de massas tornou-se uma ferramenta valiosa no estudo das proteínas.

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

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

1.10.1 Avanços Tecnológicos na Proteômica Baseada em Espectrometria de Massas

Um espectrômetro de massas tem quatro partes fundamentais: a fonte de ionização, o analisador de massas, o detector e o sistema de aquisição de dados (mostrado na **Figura**

7). Primeiramente a amostra a ser analisada é introduzida na fonte de ionização. No caso do ESI as moléculas (peptídeos) serão dissolvidas em solução polar e volátil, para então serem bombardeadas através de um fino capilar de metal (que é mantido entre 3 a 5 kV), e nebulizadas na ponta do capilar com aplicação de forte campo elétrico, permitindo a formação de pequenas gotículas carregadas. As gotículas carregadas são evaporadas por aplicação de calor e nitrogênio seco e as cargas residuais são transferidas para os peptídeos (evento denominado *droplet fission*) (121). Logo em seguida, os peptídeos ionizados são conduzidos através de campo elétrico até o analisador de massas. É possível manipular a fragmentação aumentando a tensão, e consequentemente aumentando a colisão com as moléculas de nitrogênio (122).

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

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

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

1.10.2 Fragmentação de Peptídeos

Na proteômica utiliza a espectrometria de massas para analisar peptídeos e identificar proteínas em banco de dados. Um dos primeiros métodos desenvolvidos, que usa a informação da massa molecular dos peptídeos (digeridos enzimaticamente), foi o PMF (*Peptide Mass Fingerprint*). Com o surgimento dos analisadores híbridos (dois analisadores acoplados), permitiu-se o surgimento de experimentos em sequência (*tandem*), sendo então possível detectar um determinado íon e submetê-lo a uma etapa de fragmentação (101, 132).

Na etapa de fragmentação um método que é bastante utilizado é a dissociação induzida por colisão (CID)(133). O processo de fragmentação começa com os íons sendo introduzidos na região do vácuo do espectrômetro de massas (por meio de ESI ou MALDI). Em seguida, os íons precursores são acelerados seguidos de colisões em gás neutro (hélio,

argônio ou nitrogênio), levando a conversão de energia cinética em energia vibracional. Conforme a energia vibracional excede um limiar, as ligações peptídicas covalentes podem quebrar. Os íons fragmentos resultantes são usados para obter a sequência peptídica em busca em banco de dados.

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

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

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

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

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

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

Dependendo do objetivo do trabalho, espectrometria sequencial oferece quatro tipos de análises: varredura de íons produtos, varredura de íons precursores, varredura de perda neutra e monitoramento seletivo de reações. Nesse trabalho foi utilizado um espectrômetro de massas LTQ-Orbitrap Velos com aquisição dependente de dados (DDA) (**Figura 9**). Esse tipo de análise é baseado na análise de varredura de íons precursores (*product ion scanning*): o primeiro analisador seleciona o íon de interesse (MS^1) que serão fragmentados na câmara de colisão, e os íons fragmentos serão analisados no segundo analisador de massas (MS^2) (146, 147).

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

1.10.3 Análise de Dados Proteômicos

Após todo o procedimento de preparação da amostra, processamento no espectrômetro de massas e obtenção dos espectros MS/MS , a próxima etapa é realizar a busca em banco de dados de proteínas. Inicialmente, é necessário realizar um pré-

processamento e para isso existem métodos que classificam espectros MS/MS em bons ou ruins, levando em consideração a boa relação sinal-ruído, fragmentação uniforme e extração precisa das massas de cada pico no espectro de massas. Para isso, utilizam-se algoritmos que são responsáveis pela classificação dos espectros antes de submetê-los à busca em banco de dados (148).

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

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

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

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

1.11 Biblioteca Espectral

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

O conceito de uso da biblioteca espectral inicia-se por volta da década de 70, quando a estrutura de uma espécie química era inferida pelo seu padrão de fragmentação gerado

pelo espectrômetro de massas (153). Esse método envolvia curagem manual e era demorada.

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

No âmbito da proteômica, uma idéia semelhante de busca espectral foi apliado por Yates et al em 1998 (155, 156). Yates et al obtiveram sucesso no uso de uma “correlação cruzada” normalizada para medir a semelhança entre um espectro CID e um espectro de referência obtido do mesmo instrumento. Mesmo que essa ideia tenha sido revelada bem a frente do seu tempo, bibliotecas espectrais de peptídeos não existiam antes de 1998, lançando bases para as buscas em bibliotecas espectrais como conhecemos hoje (157).

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

A busca espectral é uma forma de agregar conhecimento de experimentos anteriores em novos estudos, já que normalmente esses dados gerados ficariam sem uso.

Assim, a busca espectral é ideal para a proteômica direcionada, isto é, quando não se buscam novas informações e sim informações repetitivas de um certo conjunto pré-definido de peptídeos (152, 165). Os softwares mais usados para busca espectral são o SpectraST (*Spectra Search Tool*) (152, 166), X!Hunter (158) e Bibliospec (156), (**Figura 10 e 11**).

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

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

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

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

(157). Se o objetivo do experimento não é descobrir peptídeos novos, mas sim identificar o mesmo peptídeo em diferentes amostras, a busca por biblioteca espectral é mais adequada do que busca por banco de dados de sequências. E por fim, as limitações mencionadas na busca por biblioteca espectral podem ser parcialmente superadas pelo aumento da biblioteca espectral e também pela utilização de métodos de previsão de espectros.

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

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

A etapa final é a validação estatística, ou seja, atribuir uma probabilidade à identificação, que é considerada a combinação espectral mais próxima (157).

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

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

1.12 Proteômica no *Trypanosoma cruzi*

Devido a capacidade da proteômica baseada em espectrometria de massas em identificar e quantificar proteínas e também suas PTMs (104), é interessante aplica-la para

o estudo do *T. cruzi* já que seus diferentes estágios de desenvolvimento provavelmente têm expressão diferencial de proteínas.

Seguindo essa linha de raciocínio, o primeiro estudo realizado em *T. cruzi* utilizando LC-MS/MS foi feito em 2004 por Paba et al. (175). Neste estudo foram identificadas 1573 proteínas de 1710 espectros MS/MS. Uma análise mais rigorosa resultou na identificação de 41 proteínas com no mínimo 2 peptídeos por proteína. Das proteínas identificadas, 7 estavam relacionadas a organização celular, 7 ao metabolismo e o restante pertenciam a destinos proteicos. Além disso, 29 proteínas apresentaram expressão conservada e 9 proteínas foram mais expressas em tripomastigotas e 3 em amastigotas (175). A maioria das proteínas encontradas pertenciam a classe de proteínas de choque térmico (*Heat-Shock Proteins*, HSPs). Para validar os resultados obtidos deste estudo e de um estudo anterior, na qual foi utilizada a metodologia gel 2DE (176), Paba et al., utilizando metodologia ICAT (*Isotope-Code Affinity Tag Technology*, ICAT), encontraram taxas similares entre as duas metodologias.

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

Com o objetivo de caracterizar a expressão de proteínas na metaciclogênese, Parodi-Talice et al. (178) utilizaram géis 2DE e MALDI-TOF. Com a utilização dos géis 2DE foram obtidos 500 spots de géis, das quais foram selecionados 100 spots para análises em MALDI-TOF. Dos 100 spots de géis, 20 spots apresentaram baixa qualidade dos espectros de massas. Os 80 spots restantes apresentaram ótima qualidade, sendo que 66 spots (82%)

foram identificados em buscas em bancos de dados correspondendo a 43 proteínas diferentes. Além disso, as proteínas foram agrupadas em categorias funcionais, sendo 33% relacionadas com atividades metabólicas, 16% relacionadas com transporte de elétrons e atividades antioxidantes, 16% com biossíntese, proteólise e catabolismo, 12% estavam relacionados com proteínas estruturais, 14% foram identificadas como proteínas hipotéticas, os restantes 9% reuniam proteínas como as calmodulinas, adenilato ciclase e proteínas de superfície como a glicoproteína GP90. Em particular, os autores sugerem que as PTMs durante a metaciclogênese podem ser de grande importância na regulação da expressão gênica durante a diferenciação e que a glutamato desidrogenase (GluDH) pode desempenhar um papel importante durante o estresse oxidativo por aumentar a captação de aminoácidos como fonte de nutrientes.

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

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

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

Em um estudo para analisar a fração organelar da forma epimastigotas do *T. cruzi* conduzido por Ferrel et. al. (183) baseando-se na metodologia de Scott e Docampo (184) com utilização da LC-MS/MS, os autores identificaram 396 proteínas, das quais 138 foram anotadas como hipotéticas. Uma análise comparativa com um estudo de proteoma de células inteiras resultou na validação de 173 novas proteínas, sendo 38 proteínas não encontradas em estudos anteriores em epimastigotas. A conclusão a que os autores chegaram é que o enriquecimento de frações subcelulares pode ajudar na detecção de novas proteínas que não são encontradas em estudo proteômico em larga escala. Além da fração organelar, um outro estudo foi realizado com a fração nuclear de células de epimastigotas de *T. cruzi*, em que Santos et. al. (185) adaptaram duas metodologias (186, 187) para isolar a fração nuclear do parasito. As análises por 2D-nLC-MS/MS (*Two-Dimensional nano-Scale Liquid Chromatography*) revelaram 12875 peptídeos de 864 proteínas. Esse foi o primeiro estudo proteômico em larga escala realizado de fração nuclear. Dentre as 864 proteínas, 272 foram anotadas como não caracterizadas, e 275 não foram

encontradas em análise global do proteoma do *T. cruzi*. Para confirmar o método desenvolvido, análises de bioinformática foram realizadas, revelando 65 cluster de genes, em que foi mostrado o agrupamento de membros com organização de cromatina e funções ligadas ao DNA.

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

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

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

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

6. CONCLUSÃO

Neste estudo, apresentamos uma caracterização de DTUs de *T. cruzi* com base na correspondência espectral MS/MS. O método Tc-STAMS² é robusto, sensível e poderoso com base na identificação pelo espectro MS/MS mostrando ser um método de caracterização para complementar outros já estabelecidos. O método proposto pode ajudar pesquisadores na caracterização das DTUs sem a necessidade de dados genômicos, o que é muito escasso para esse parasito.

REFERÊNCIAS*

*De acordo com:

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

1. Chagas C. Nova tripanozomíase humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. Memórias do Instituto Oswaldo Cruz. 1909;1:159-218.
2. WHO. Chagas disease (American trypanosomiasis): World Health Organization 2017 [Available from: <http://www.who.int/mediacentre/factsheets/fs340/en/>].
3. PAHO. Chagas disease. Pan American Health Organization; 2017.
4. Rassi A, Jr., Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010;375(9723):1388-402.
5. Victora CG, Barreto ML, do Carmo Leal M, Monteiro CA, Schmidt MI, Paim J, et al. Health conditions and health-policy innovations in Brazil: the way forward. Lancet. 2011;377(9782):2042-53.
6. Martins-Melo FR, Ramos AN, Jr., Alencar CH, Heukelbach J. Prevalence of Chagas disease in Brazil: a systematic review and meta-analysis. Acta Trop. 2014;130:167-74.
7. Hotez PJ, Fujiwara RT. Brazil's neglected tropical diseases: an overview and a report card. Microbes Infect. 2014;16(8):601-6.
8. Dias JC, Ramos AN, Jr., Gontijo ED, Luquetti A, Shikanai-Yasuda MA, Coura JR, et al. [Brazilian Consensus on Chagas Disease, 2015]. Epidemiol Serv Saude. 2016;25(spe):7-86.
9. Dias JCP, Coura JR, Yasuda MAS. The present situation, challenges, and perspectives regarding the production and utilization of effective drugs against human Chagas disease. Revista da Sociedade Brasileira de Medicina Tropical. 2014;47:123-5.
10. Dias J, Silveira A, Schofield C. The impact of Chagas disease control in Latin America: a review. Memórias do Instituto Oswaldo Cruz. 2002;97:603-12.
11. Barreto ML, Teixeira MG, Bastos FI, Ximenes RA, Barata RB, Rodrigues LC. Successes and failures in the control of infectious diseases in Brazil: social and environmental

- context, policies, interventions, and research needs. *Lancet*. 2011;377(9780):1877-89.
12. Martins-Melo FR, Ramos Junior AN, Alencar CH, Heukelbach J. Multiple causes of death related to Chagas' disease in Brazil, 1999 to 2007. *Revista da Sociedade Brasileira de Medicina Tropical*. 2012;45:591-6.
 13. Martins-Melo FR, Ramos AN, Jr., Alencar CH, Heukelbach J. Mortality due to Chagas disease in Brazil from 1979 to 2009: trends and regional differences. *J Infect Dev Ctries*. 2012;6(11):817-24.
 14. da Nobrega AA, de Araujo WN, Vasconcelos AM. Mortality due to Chagas disease in Brazil according to a specific cause. *Am J Trop Med Hyg*. 2014;91(3):528-33.
 15. Saúde Md. Boletim Epidemiológico. Ministério da Saúde - Secretaria de Vigilância em Saúde 2015.
 16. Cavalier-Smith T. Eukaryote kingdoms: Seven or nine? *Biosystems*. 1981;14(3):461-81.
 17. Vickerman K. The diversity of the kinetoplastid flagellates. In *Biology of the Kinetoplastida*. Evans WHRLDA, editor: Academic Press; 1976.
 18. Liu B, Liu Y, Motyka SA, Agbo EE, Englund PT. Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol*. 2005;21(8):363-9.
 19. Moreira D, Lopez-Garcia P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol*. 2004;54(Pt 5):1861-75.
 20. Simpson AGB, Roger AJ. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Molecular Phylogenetics and Evolution*. 2004;30(1):201-12.
 21. Jankevicius JV, Itow-Jankevicius S, Maeda LA, Campaner M, Conchon I, do Carmo JB, et al. [Biological cycle of *Phytomonas*]. *Mem Inst Oswaldo Cruz*. 1988;83 Suppl 1:601-10.
 22. Leonard G, Soanes DM, Stevens JR. Resolving the question of trypanosome monophyly: a comparative genomics approach using whole genome data sets with low taxon sampling. *Infect Genet Evol*. 2011;11(5):955-9.
 23. Maslov DA, Votycka J, Yurchenko V, Lukes J. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol*. 2013;29(1):43-52.
 24. Lukes J, Skalicky T, Tyc J, Votycka J, Yurchenko V. Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol*. 2014;195(2):115-22.
 25. Porcel BM, Denoeud F, Opperdoes F, Noel B, Madoui MA, Hammarton TC, et al. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet*. 2014;10(2):e1004007.
 26. Levine ND. **The Trypanosomes of Mammals.** A Zoological Monograph. Cecil A. Hoare. Blackwell, Oxford, England, 1972 (U.S. distributor, Davis, Philadelphia). xviii, 750 pp. + plates. \$34.50. *Science*. 1973;179(4068):60-.

27. A Podlipaev S, Frolov A. [The phylogeny of the Trypanosomatidae: the molecular and morphological approaches]2000. 169-82 p.
28. Bonaldo MC, Souto-Padron T, de Souza W, Goldenberg S. Cell-substrate adhesion during *Trypanosoma cruzi* differentiation. *J Cell Biol.* 1988;106(4):1349-58.
29. Kleffmann T, Schmidt J, Schaub GA. Attachment of *Trypanosoma cruzi* epimastigotes to hydrophobic substrates and use of this property to separate stages and promote metacyclogenesis. *J Eukaryot Microbiol.* 1998;45(5):548-55.
30. Tyler KM, Engman DM. The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol.* 2001;31(5-6):472-81.
31. Brener Z. *Trypanosoma cruzi*: morfologia e ciclo evolutivo. In: João Carlos Pinto Dias JRC, editor. *Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral.* Scielo Books: FIOCRUZ; 1997. p. 486.
32. Rodriguez A, Webster P, Ortego J, Andrews NW. Lysosomes behave as Ca²⁺-regulated exocytic vesicles in fibroblasts and epithelial cells. *J Cell Biol.* 1997;137(1):93-104.
33. Andrews NW. Living dangerously: how *Trypanosoma cruzi* uses lysosomes to get inside host cells, and then escapes into the cytoplasm. *Biol Res.* 1993;26(1-2):65-7.
34. CDC. Parasites - American Trypanosomiasis (also known as Chagas Disease) Center for Disease Control and Prevention2017 [Available from: <https://www.cdc.gov/parasites/chagas/biology.html>].
35. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz.* 2009;104(7):1051-4.
36. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol.* 2012;12(2):240-53.
37. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology.* 2009;136(6):641-55.
38. Lima L, Espinosa-Alvarez O, Ortiz PA, Trejo-Varon JA, Carranza JC, Pinto CM, et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Trop.* 2015;151:166-77.
39. Andrade SG. Caracterização de cepas do *Trypanosoma cruzi* isoladas no Recôncavo Baiano. *Revista Patologia Tropical.* 31974. p. 65-121.
40. Magalhães JB, Andrade SG. Estudo do comportamento de cepas de *Trypanosoma cruzi* após passagem em diferentes espécies de triatomíneos. *Revista da Sociedade Brasileira de Medicina Tropical.* 1991;24:209-16.
41. Filardi LS, Brener Z. A rapid method for testing in vivo the susceptibility of different strains of *Trypanosoma cruzi* to active chemotherapeutic agents. *Mem Inst Oswaldo Cruz.* 1984;79(2):221-5.
42. Miles MA, Tøye PJ, Oswald SC, Godfrey DG. The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Trans R Soc Trop Med Hyg.* 1977;71(3):217-25.

43. Tibayrenc M, Ayala FJ. Isozyme Variability in *Trypanosoma Cruzi*, the Agent of Chagas' Disease: Genetical, Taxonomical, and Epidemiological Significance. *Evolution*. 1988;42(2):277-92.
44. Miles MA, Souza A, Povia M, Shaw JJ, Lainson R, Toye PJ. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature*. 1978;272(5656):819-21.
45. Miles MA, Lanham SM, de Souza AA, Povia M. Further enzymic characters of *Trypanosoma cruzi* and their evaluation for strain identification. *Trans R Soc Trop Med Hyg*. 1980;74(2):221-37.
46. Tibayrenc M, Ward P, Moya A, Ayala FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proc Natl Acad Sci U S A*. 1986;83(1):115-9.
47. Tibayrenc M, Cariou ML, Solignac M, Carlier Y. Genetic arguments against actual mating in *Trypanosoma cruzi*. Taxonomic implications. *Comptes Rendus des Seances de l'Academie des Sciences - Series III*. 1981;293(3):207-9.
48. Steindel M, Dias Neto E, de Menezes CL, Romanha AJ, Simpson AJ. Random amplified polymorphic DNA analysis of *Trypanosoma cruzi* strains. *Mol Biochem Parasitol*. 1993;60(1):71-9.
49. Morel C, Chiari E, Camargo EP, Mattei DM, Romanha AJ, Simpson L. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proc Natl Acad Sci U S A*. 1980;77(11):6810-4.
50. Henriksson J, Aslund L, Macina RA, Franke de Cazzulo BM, Cazzulo JJ, Frasch AC, et al. Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi*. *Mol Biochem Parasitol*. 1990;42(2):213-23.
51. Macedo AM, Melo MN, Gomes RF, Pena SD. DNA fingerprints: a tool for identification and determination of the relationships between species and strains of *Leishmania*. *Mol Biochem Parasitol*. 1992;53(1-2):63-70.
52. Brisse S, Barnabe C, Banuls AL, Sidibe I, Noel S, Tibayrenc M. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. *Mol Biochem Parasitol*. 1998;92(2):253-63.
53. Sturm NR, Vargas NS, Westenberger SJ, Zingales B, Campbell DA. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *Int J Parasitol*. 2003;33(3):269-79.
54. Westenberger SJ, Barnabe C, Campbell DA, Sturm NR. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics*. 2005;171(2):527-43.
55. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *Int J Parasitol*. 2009;39(12):1305-17.
56. Ochaya S. Comparative genomics and molecular characterization of N-alpha Acetyltransferase in *Trypanosomes* for drug target identification: Karolinska Institutet; 2013.

57. Macedo AM, Machado CR, Oliveira RP, Pena SD. *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease. *Mem Inst Oswaldo Cruz*. 2004;99(1):1-12.
58. Zingales B. *Trypanosoma cruzi*: um parasita, dois parasitas ou vários parasitas da doença de chagas? 2011. 44-8 p.
59. Cosentino RO, Aguero F. A simple strain typing assay for *Trypanosoma cruzi*: discrimination of major evolutionary lineages from a single amplification product. *PLoS Negl Trop Dis*. 2012;6(7):e1777.
60. Tibayrenc M, Neubauer K, Barnabe C, Guerrini F, Skarecky D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proc Natl Acad Sci U S A*. 1993;90(4):1335-9.
61. Brisse S, Verhoef J, Tibayrenc M. Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *Int J Parasitol*. 2001;31(11):1218-26.
62. Cura CI, Duffy T, Lucero RH, Bisio M, Peneau J, Jimenez-Coello M, et al. Multiplex Real-Time PCR Assay Using TaqMan Probes for the Identification of *Trypanosoma cruzi* DTUs in Biological and Clinical Samples. *PLoS Negl Trop Dis*. 2015;9(5):e0003765.
63. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94(3):441-8.
64. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 2005;309(5733):409-15.
65. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS Negl Trop Dis*. 2011;5(3):e984.
66. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, Segovia M, et al. *Trypanosoma cruzi* Ilc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. *PLoS Negl Trop Dis*. 2009;3(9):e510.
67. Grisard EC, Teixeira SM, de Almeida LG, Stoco PH, Gerber AL, Talavera-Lopez C, et al. *Trypanosoma cruzi* Clone Dm28c Draft Genome Sequence. *Genome Announc*. 2014;2(1).
68. Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, Lewis MD, et al. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC Genomics*. 2012;13:531.
69. Bartholomeu DC, de Paiva RM, Mendes TA, DaRocha WD, Teixeira SM. Unveiling the intracellular survival gene kit of trypanosomatid parasites. *PLoS Pathog*. 2014;10(12):e1004399.
70. Kramer S, Carrington M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol*. 2011;27(1):23-30.

71. Johnson PJ, Kooter JM, Borst P. Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. *Cell*. 1987;51(2):273-81.
72. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell*. 2003;11(5):1291-9.
73. De Gaudenzi JG, Noe G, Campo VA, Frasch AC, Cassola A. Gene expression regulation in trypanosomatids. *Essays Biochem*. 2011;51:31-46.
74. Agabian N. Trans splicing of nuclear pre-mRNAs. *Cell*. 1990;61(7):1157-60.
75. Preusser C, Jae N, Bindereif A. mRNA splicing in trypanosomes. *Int J Med Microbiol*. 2012;302(4-5):221-4.
76. Stuart K, Allen TE, Heidmann S, Seiwert SD. RNA editing in kinetoplastid protozoa. *Microbiol Mol Biol Rev*. 1997;61(1):105-20.
77. Hajduk S, Ochsenreiter T. RNA editing in kinetoplastids. *RNA Biol*. 2010;7(2):229-36.
78. Rudenko G, Chung HM, Pham VP, Van der Ploeg LH. RNA polymerase I can mediate expression of CAT and neo protein-coding genes in *Trypanosoma brucei*. *EMBO J*. 1991;10(11):3387-97.
79. Lee MG, Van der Ploeg LH. Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol*. 1997;51:463-89.
80. Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walte. *Molecular Biology of the Cell*: Garland Science; 2002.
81. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev*. 2007;21(15):1833-56.
82. Zomerdijs JC, Kieft R, Borst P. Efficient production of functional mRNA mediated by RNA polymerase I in *Trypanosoma brucei*. *Nature*. 1991;353(6346):772-5.
83. Rudenko G, Blundell PA, Dirks-Mulder A, Kieft R, Borst P. A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in *T. brucei*. *Cell*. 1995;83(4):547-53.
84. Yan S, Lodes MJ, Fox M, Myler PJ, Stuart K. Characterization of the *Leishmania donovani* ribosomal RNA promoter. *Mol Biochem Parasitol*. 1999;103(2):197-210.
85. McAndrew M, Graham S, Hartmann C, Clayton C. Testing Promoter Activity in the Trypanosome Genome: Isolation of a Metacyclic-Type VSG Promoter, and Unexpected Insights into RNA Polymerase II Transcription. *Experimental Parasitology*. 1998;90(1):65-76.
86. Parsons M, Nelson RG, Watkins KP, Agabian N. Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell*. 1984;38(1):309-16.
87. Araujo PR, Teixeira SM. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. *Mem Inst Oswaldo Cruz*. 2011;106(3):257-66.
88. Dias JC, Coura J. *Clínica e Terapêutica da Doença de Chagas: uma abordagem prática para o clínico geral*. João Carlos Pinto Dias JRC, editor. FIOCRUZ: FIOCRUZ; 1997.
89. Antas PR, Medrano-Mercado N, Torrico F, Ugarte-Fernandez R, Gomez F, Correa Oliveira R, et al. Early, intermediate, and late acute stages in Chagas' disease: a study

- combining anti-galactose IgG, specific serodiagnosis, and polymerase chain reaction analysis. *Am J Trop Med Hyg.* 1999;61(2):308-14.
90. Andrade DV, Gollob KJ, Dutra WO. Acute chagas disease: new global challenges for an old neglected disease. *PLoS Negl Trop Dis.* 2014;8(7):e3010.
 91. Punukollu G, Gowda RM, Khan IA, Navarro VS, Vasavada BC. Clinical aspects of the Chagas' heart disease. *Int J Cardiol.* 2007;115(3):279-83.
 92. Linardi DPNALDMPM. *Parasitologia Humana*: Atheneu; 2005.
 93. Prata A. Clinical and epidemiological aspects of Chagas disease. *Lancet Infect Dis.* 2001;1(2):92-100.
 94. Docampo R. Sensitivity of parasites to free radical damage by antiparasitic drugs. *Chem Biol Interact.* 1990;73(1):1-27.
 95. Rodrigues Coura J, de Castro SL. A critical review on Chagas disease chemotherapy. *Mem Inst Oswaldo Cruz.* 2002;97(1):3-24.
 96. Bern C, Montgomery SP, Herwaldt BL, Rassi A, Jr., Marin-Neto JA, Dantas RO, et al. Evaluation and treatment of chagas disease in the United States: a systematic review. *JAMA.* 2007;298(18):2171-81.
 97. Gomes ML, Galvao LM, Macedo AM, Pena SD, Chiari E. Chagas' disease diagnosis: comparative analysis of parasitologic, molecular, and serologic methods. *Am J Trop Med Hyg.* 1999;60(2):205-10.
 98. Carlier Y, Torrico F, Sosa-Estani S, Russomando G, Luquetti A, Freilij H, et al. Congenital Chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women. *PLoS Negl Trop Dis.* 2011;5(10):e1250.
 99. Moreira LM. *Ciências Genômicas: Fundamentos e Aplicações*: Editora Cubo; 2015.
 100. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis.* 1995;16(7):1090-4.
 101. Yates JR, 3rd. Mass spectrometry and the age of the proteome. *J Mass Spectrom.* 1998;33(1):1-19.
 102. de Hoog CL, Mann M. Proteomics. *Annu Rev Genomics Hum Genet.* 2004;5:267-93.
 103. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63.
 104. Bensimon A, Heck AJ, Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem.* 2012;81:379-405.
 105. Anderson NL, Hofmann JP, Gemmell A, Taylor J. Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis. *Clin Chem.* 1984;30(12 Pt 1):2031-6.
 106. Tarroux P, Vincens P, Rabilloud T. HERMeS: A second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part V: Data analysis. *ELECTROPHORESIS.* 1987.
 107. Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics.* 2009;2009.
 108. Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem.* 2001;73(23):5683-90.

109. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol.* 1999;17(7):676-82.
110. Rabilloud T, Chevallet M, Luche S, Lelong C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics.* 2010;73(11):2064-77.
111. Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF. Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. *Electrophoresis.* 1998;19(8-9):1501-5.
112. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis.* 2000;21(6):1104-15.
113. Rabilloud T. Membrane proteins and proteomics: love is possible, but so difficult. *Electrophoresis.* 2009;30 Suppl 1:S174-80.
114. Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. *J Chromatogr A.* 2010;1217(25):3938-45.
115. Hilton GR, Benesch JL. Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J R Soc Interface.* 2012;9(70):801-16.
116. Mallick P, Kuster B. Proteomics: a pragmatic perspective. *Nat Biotechnol.* 2010;28(7):695-709.
117. Barbosa EB, Vidotto A, Polachini GM, Henrique T, Trovó de Marqui AB, Tajara EH. Proteômica: metodologias e aplicações no estudo de doenças humanas. *Revista da Associação Médica Brasileira.* 2012;58(3):366-75.
118. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet.* 2013;14(1):35-48.
119. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat Genet.* 2003;33 Suppl:311-23.
120. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol.* 2004;5(9):699-711.
121. Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom.* 2000;35(7):804-17.
122. Carr SA, Huddleston MJ, Bean MF. Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci.* 1993;2(2):183-96.
123. Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov.* 2003;2(2):140-50.
124. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science.* 2006;312(5771):212-7.
125. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003;422(6928):198-207.
126. Lanças FM. A cromatografia líquida moderna e a espectrometria de massas: Finalmente “compatíveis”? II. A escolha do analisador de massas. *Scientia Chromatographica.* 2013.
127. Das A. Spinning Charged Test-Particles in General Relativity. *Progress of Theoretical Physics.* 1960;23(4):610-5.

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

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

163. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, et al. PRIDE: the proteomics identifications database. *Proteomics*. 2005;5(13):3537-45.
164. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res*. 2006;34(Database issue):D655-8.
165. Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol*. 2005;6(7):577-83.
166. Lam H, Deutsch EW, Eddes JS, Eng JK, Stein SE, Aebersold R. Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods*. 2008;5(10):873-5.
167. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005;1:2005 0017.
168. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry*. 1994;5(9):859-66.
169. Pedrioli PG. Trans-proteomic pipeline: a pipeline for proteomic analysis. *Methods Mol Biol*. 2010;604:213-38.
170. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics*. 2010;73(11):2092-123.
171. Lam H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics*. 2011;10(12):R111 008565.
172. Lam H, Aebersold R. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods*. 2011;54(4):424-31.
173. Onder O, Shao W, Kemps BD, Lam H, Brisson D. Identifying sources of tick blood meals using unidentified tandem mass spectral libraries. *Nat Commun*. 2013;4:1746.
174. Shao W, Zhang M, Lam H, Lau SC. A peptide identification-free, genome sequence-independent shotgun proteomics workflow for strain-level bacterial differentiation. *Sci Rep*. 2015;5:14337.
175. Paba J, Ricart CA, Fontes W, Santana JM, Teixeira AR, Marchese J, et al. Proteomic analysis of *Trypanosoma cruzi* developmental stages using isotope-coded affinity tag reagents. *J Proteome Res*. 2004;3(3):517-24.
176. Paba J, Santana JM, Teixeira AR, Fontes W, Sousa MV, Ricart CA. Proteomic analysis of the human pathogen *Trypanosoma cruzi*. *Proteomics*. 2004;4(4):1052-9.
177. Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, et al. The *Trypanosoma cruzi* proteome. *Science*. 2005;309(5733):473-6.
178. Parodi-Talice A, Monteiro-Goes V, Arrambide N, Avila AR, Duran R, Correa A, et al. Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. *J Mass Spectrom*. 2007;42(11):1422-32.
179. Telleria J, Biron DG, Brizard JP, Demetree E, Seveno M, Barnabe C, et al. Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*. *Proc Natl Acad Sci U S A*. 2010;107(47):20411-6.

180. Diaz ML, Torres R, Gonzalez CI. [Differential protein expression in developmental stages of *Trypanosoma cruzi* I isolated from a patient with chronic chagasic cardiomyopathy]. *Biomedica*. 2011;31(4):503-13.
181. Queiroz RM, Charneau S, Motta FN, Santana JM, Roepstorff P, Ricart CA. Comprehensive proteomic analysis of *Trypanosoma cruzi* epimastigote cell surface proteins by two complementary methods. *J Proteome Res*. 2013;12(7):3255-63.
182. Queiroz RM, Charneau S, Bastos IM, Santana JM, Sousa MV, Roepstorff P, et al. Cell surface proteome analysis of human-hosted *Trypanosoma cruzi* life stages. *J Proteome Res*. 2014;13(8):3530-41.
183. Ferella M, Nilsson D, Darban H, Rodrigues C, Bontempi EJ, Docampo R, et al. Proteomics in *Trypanosoma cruzi*--localization of novel proteins to various organelles. *Proteomics*. 2008;8(13):2735-49.
184. Scott DA, Docampo R. Characterization of isolated acidocalcisomes of *Trypanosoma cruzi*. *J Biol Chem*. 2000;275(31):24215-21.
185. dos Santos Junior Ade C, Kalume DE, Camargo R, Gomez-Mendoza DP, Correa JR, Charneau S, et al. Unveiling the *Trypanosoma cruzi* Nuclear Proteome. *PLoS One*. 2015;10(9):e0138667.
186. Batista JA, Teixeira SM, Donelson JE, Kirchhoff LV, de Sa CM. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Mol Biochem Parasitol*. 1994;67(2):301-12.
187. Rout MP, Field MC. Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. *J Biol Chem*. 2001;276(41):38261-71.
188. Miles MA, Cedillos RA, Povia MM, de Souza AA, Prata A, Macedo V. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? *Lancet*. 1981;1(8234):1338-40.
189. Camargo EP. Growth and Differentiation in *Trypanosoma Cruzi*. I. Origin of Metacyclic Trypanosomes in Liquid Media. *Rev Inst Med Trop Sao Paulo*. 1964;6:93-100.
190. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*. 2007;2(8):1896-906.
191. Palmisano G, Parker BL, Engholm-Keller K, Lendal SE, Kulej K, Schulz M, et al. A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Mol Cell Proteomics*. 2012;11(11):1191-202.
192. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008;24(21):2534-6.
193. Efron B, Robert. *An Introduction to the Bootstrap* (Chapman & Hall/CRC Monographs on Statistics & Applied Probability): Chapman and Hall/CRC; 1994.
194. Silva ARF, Lima DB, Leyva A, Duran R, Batthyany C, Aquino PF, et al. DiagnoProt: a tool for discovery of new molecules by mass spectrometry. *Bioinformatics*. 2017;33(12):1883-5.

195. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics*. 2013;13(1):22-4.
196. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002;74(20):5383-92.
197. Lee HJ, Jeong SK, Na K, Lee MJ, Lee SH, Lim JS, et al. Comprehensive genome-wide proteomic analysis of human placental tissue for the Chromosome-Centric Human Proteome Project. *J Proteome Res*. 2013;12(6):2458-66.
198. Jackson AP, Goyard S, Xia D, Foth BJ, Sanders M, Wastling JM, et al. Global Gene Expression Profiling through the Complete Life Cycle of *Trypanosoma vivax*. *PLoS Negl Trop Dis*. 2015;9(8):e0003975.
199. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447-52.
200. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731-40.
201. Jabbour RE, Deshpande SV, Wade MM, Stanford MF, Wick CH, Zulich AW, et al. Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. *Appl Environ Microbiol*. 2010;76(11):3637-44.
202. Dworzanski JP, Snyder AP, Chen R, Zhang H, Wishart D, Li L. Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring. *Anal Chem*. 2004;76(8):2355-66.
203. Dworzanski JP, Deshpande SV, Chen R, Jabbour RE, Snyder AP, Wick CH, et al. Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *J Proteome Res*. 2006;5(1):76-87.
204. Baker JR, Miles MA, Godfrey DG, Barrett TV. Biochemical characterization of some species of *Trypanosoma* (*Schizotrypanum*) from bats (*Microchiroptera*). *Am J Trop Med Hyg*. 1978;27(3):483-91.
205. Ebert F. Comparison of isoenzymes of some species of the subgenus *schizotrypanum* from bats by isoelectrofocusing. *Tropenmed Parasitol*. 1983;34(2):93-7.
206. Tibayrenc M, Le Ray D. General classification of the isoenzymic strains of *Trypanosoma* (*Schizotrypanum*) *cruzi* and comparison with *T. (S.) C. marinkellei* and *T. (Herpetosoma) rangeli*. *Ann Soc Belg Med Trop*. 1984;64(3):239-48.
207. Petry K, Baltz T, Schottelius J. Differentiation of *Trypanosoma cruzi*, *T. cruzi marinkellei*, *T. dionisii* and *T. vespertilionis* by monoclonal antibodies. *Acta Trop*. 1986;43(1):5-13.
208. Petry K, Voisin P, Baltz T. Complex lipids as common antigens to *Trypanosoma cruzi*, *T. dionisii*, *T. vespertilionis* and nervous tissue (astrocytes, neurons). *Acta Trop*. 1987;44(4):381-6.
209. Stevens JR, Noyes HA, Schofield CJ, Gibson W. The molecular evolution of *Trypanosomatidae*. *Adv Parasitol*. 2001;48:1-56.
210. Barnabe C, Brisse S, Tibayrenc M. Phylogenetic diversity of bat trypanosomes of subgenus *Schizotrypanum* based on multilocus enzyme electrophoresis, random

- amplified polymorphic DNA, and cytochrome b nucleotide sequence analyses. *Infect Genet Evol.* 2003;2(3):201-8.
211. Telleria J, Barnabe C, Hide M, Banuls AL, Tibayrenc M. Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi*. *Mol Biochem Parasitol.* 2004;137(1):133-41.
 212. Lima L, Ortiz PA, da Silva FM, Alves JM, Serrano MG, Cortez AP, et al. Repertoire, genealogy and genomic organization of cruzipain and homologous genes in *Trypanosoma cruzi*, *T. cruzi*-like and other trypanosome species. *PLoS One.* 2012;7(6):e38385.
 213. Brener Z, Chiari E. Variações Morfológicas Observadas em Diferentes Amostras de *Trypanosoma cruzi*. *Rev Inst Med Trop Sao Paulo.* 1963.
 214. Melo RC, Brener Z. Tissue tropism of different *Trypanosoma cruzi* strains. *J Parasitol.* 1978;64(3):475-82.
 215. Atayde VD, Neira I, Cortez M, Ferreira D, Freymuller E, Yoshida N. Molecular basis of non-virulence of *Trypanosoma cruzi* clone CL-14. *Int J Parasitol.* 2004;34(7):851-60.
 216. Yoshida N. Surface antigens of metacyclic trypomastigotes of *Trypanosoma cruzi*. *Infect Immun.* 1983;40(2):836-9.
 217. Fairlamb AH, Cerami A. Metabolism and functions of trypanothione in the Kinetoplastida. *Annu Rev Microbiol.* 1992;46:695-729.
 218. Piacenza L, Zago MP, Peluffo G, Alvarez MN, Basombrio MA, Radi R. Enzymes of the antioxidant network as novel determiners of *Trypanosoma cruzi* virulence. *Int J Parasitol.* 2009;39(13):1455-64.
 219. Zago MP, Hosakote YM, Koo S-j, Dhiman M, Piñeyro MD, Parodi-Talice A, et al. Tci Isolates of *Trypanosoma cruzi* Exploit the Antioxidant Network for Enhanced Intracellular Survival in Macrophages and Virulence in Mice. *Infection and Immunity.* 2016;84(6):1842-56.
 220. Giese V, Dallagiovanna B, Marchini FK, Pavoni DP, Krieger MA, Goldenberg S. *Trypanosoma cruzi*: a stage-specific calpain-like protein is induced after various kinds of stress. *Mem Inst Oswaldo Cruz.* 2008;103(6):598-601.
 221. Osorio L, Rios I, Gutierrez B, Gonzalez J. Virulence factors of *Trypanosoma cruzi*: who is who? *Microbes Infect.* 2012;14(15):1390-402.
 222. Perez Brandan C, Padilla AM, Xu D, Tarleton RL, Basombrio MA. Knockout of the dhfr-ts gene in *Trypanosoma cruzi* generates attenuated parasites able to confer protection against a virulent challenge. *PLoS Negl Trop Dis.* 2011;5(12):e1418.
 223. Avila CC, Almeida FG, Palmisano G. Direct identification of trypanosomatids by matrix-assisted laser desorption ionization-time of flight mass spectrometry (DIT MALDI-TOF MS). *J Mass Spectrom.* 2016;51(8):549-57.