



# Universidade de Brasília

Faculdade de Medicina

Programa de Pós-Graduação em Patologia Molecular

Laboratório de Interação Patógeno-Hospedeiro

## **MODELAGEM ESTRUTURAL E ANÁLISES *IN SILICO* E *IN VITRO* DA CATEPSINA B DE *TRYPANOSOMA CRUZI***

ARTHUR SOARES DE ARAUJO

Brasília - DF, 2021.

ARTHUR SOARES DE ARAUJO

**MODELAGEM ESTRUTURAL E ANÁLISES *IN SILICO* E *IN VITRO* DA  
CATEPSINA B DE *TRYPANOSOMA CRUZI***

**ARTHUR SOARES DE ARAUJO**

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Patologia Molecular da Universidade de Brasília – Faculdade de Medicina como requisito parcial para a obtenção do grau de Mestre em Patologia Molecular.

Orientadora: **Prof<sup>a</sup>. Dr<sup>a</sup>. Izabela Marques Dourado Bastos Charneau**

Coorientadora: **Prof<sup>a</sup>. Dr<sup>a</sup>. Melina Mottin**

Brasília - DF, 2021.

Este estudo foi desenvolvido no Laboratório de Interação Patógeno-Hospedeiro (LIPH) do Departamento de Biologia Celular da Universidade de Brasília.

*Apoio financeiro:* Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF, Processos 193.001.053/2015 e 193.001.802/2017).

Projeto de Mestrado realizado com bolsa da CAPES e CNPq

*“Tudo aquilo que o homem ignora, não existe para ele. Por isso, o universo de cada um, se resume no tamanho do seu saber.”*

*(Albert Einstein)*

*Dedico este trabalho aos meus pais e minha  
irmã. Sem eles eu nunca alcançaria tudo que  
eu conquistei.*

## AGRADECIMENTOS

À **Profª. Dra. Izabela Marques Dourado Bastos**, minha orientadora, por ter acreditado em mim, recém graduado e sem experiência nas técnicas desenvolvidas em sua pesquisa, mesmo assim me aceitou em seu laboratório. Assim, me possibilitou a experiência de uma pós-graduação em uma instituição de excelência atuando em uma área que sou apaixonado e as futuras oportunidades na minha carreira. Obrigado pela orientação e por dividir a sua vasta experiência profissional comigo.

À **Dra. Melina Mottin**, minha coorientadora, por toda paciência em me explicar o que eu não sabia e/ou entendia e a disponibilidade no desenvolvimento do meu projeto. A sua participação, mesmo que não seja desde o começo, enriqueceu muito meu trabalho. Seu interesse e preocupação foram essenciais para a quantidade e qualidade dos resultados *in silico* obtidos.

Aos meus pais, **Geize Soares de Oliveira Araujo** e **Antônio Wellington Araujo do Nascimento**, pelo amor, dedicação e apoio. Agradeço pelos momentos ausentes trabalhando e se esforçando sempre procurando me fornecer uma condição muito melhor do que vocês tiveram. Agradeço o incentivo e investimentos na minha educação, eu nunca alcançaria os feitos que conquistei sem este suporte. Agradeço pelos infinitos privilégios que vocês me proporcionaram, todos eles foram fundamentais para o acesso à cultura, educação, conhecimento, lazer e às múltiplas portas que sempre estiveram abertas no meu caminho e me concederam o poder de escolha do caminho que eu gostaria de trilhar.

À minha irmã, **Letícia Soares de Araujo**, por todos os momentos que a gente compartilhou desde que você nasceu, eu aprendi e cresci muito desde então. Agradeço o companheirismo e apoio quando eu precisei, principalmente durante o período de pandemia.

Às **Profas. Dras. Carla Nunes** e **Flávia Nader Motta**, por sempre estarem dispostas a oferecer ajuda e esclarecer dúvidas e dificuldades em relação ao projeto, ao **Prof. Dr. Jaime Martins de Santana** por disponibilizar acesso ao laboratório e aos recursos necessários para o desenvolvimento da minha pesquisa e ao **Prof. Philippe Grellier** do Museu Nacional de História Natural de Paris (MNHN) pela colaboração e disponibilização da biblioteca de compostos.

Aos meus colegas e amigos do LIPH, atuais e egressos, especialmente, **Agnelo Rodrigues, Alessandro Zanard, Allan Mascarenhas, Amanda Rocha, Camila Lasse, Clênia Azevedo, Isabela Cardoso, Juliana Name, Kaio Bentes, Laís Garreto, Milene Andrade, Natália Gil, Paula Santiago e Yanna Praça.** Agradeço por me acolherem tão bem no LIPH fazendo esse tempo ser muito especial e descomplicado. Agradeço por todo carinho, atenção, disposição, apoio, ajuda, conselhos, desabafos, preocupação e conversas na copa. Vocês foram fundamentais para gerar o equilíbrio necessário para confrontar as adversidades e problemas que enfrentamos tanto nas nossas pesquisas quanto na carreira dentro da ciência. Aos outros membros do LIPH, **Gabriel dos Santos, Isabela Roma, Karen Mangabeira, Marcelle Caminha, Marta Barbosa, Nathalia Araujo e Sofia Marcelino,** por toda a ajuda e disponibilidade durante meu tempo no laboratório.

Aos meus amigos de graduação e da vida, **Ana Luiza Oliveira, Maria Beatriz Nóbrega, Ian Holanda, Gabriel Carneiro e Natasha Evangelista.** Obrigado por estarem presentes nos momentos bons e ruins. Agradeço por todas as ocasiões divididas, onde sempre houve muita felicidade, amor, confiança e apoio. Espero que vocês saibam o quanto eu os considero especiais na minha vida.

Ao **Ricardo Pereira de Moura,** você me acompanhou de perto durante todo o período de mestrado, da entrada à saída, fornecendo todo o apoio e os conselhos necessários nos momentos de maiores dificuldade. A sua presença e perspectiva sobre as coisas sempre foram cruciais para me centralizar e me fazer retomar o foco ao que era importante. Obrigado por me mostrar como o amor pode ser leve.

À **Universidade de Brasília,** ambiente que foi praticamente minha segunda casa durante todos esses anos, onde pude ter acesso a uma educação gratuita e de qualidade e me proporcionou os melhores anos da minha vida. Esse espaço não me concedeu somente formação acadêmica, mas também como cidadão.

Às instituições de fomento, **CAPES, CNPq, FAPDF e FINATEC,** por financiarem esse projeto permitindo, assim, sua realização.

# SUMÁRIO

LISTA DE ABREVIATURAS.....	9
LISTA DE FIGURAS.....	10
LISTA DE TABELAS.....	15
RESUMO.....	16
ABSTRACT.....	17
1 INTRODUÇÃO.....	19
1.1 Doença de Chagas.....	19
1.2 Distribuição da doença.....	20
1.3 Ciclo biológico do <i>Trypanosoma cruzi</i> .....	22
1.4 Tratamento da doença de Chagas.....	24
1.5 Proteases.....	26
1.6 Catepsinas.....	27
1.6.1 Catepsina B.....	28
1.6.2 Catepsina B de protozoários.....	32
1.7 Genética e biologia molecular do <i>Trypanosoma cruzi</i> .....	34
1.8 Utilização do sistema CRISPR/Cas9.....	37
1.8.1 Edição gênica em <i>Trypanosoma cruzi</i> .....	39
1.9 Modelagem da estrutura tridimensional da catepsina B de <i>Trypanosoma cruzi</i> .....	40
1.10 Triagem virtual baseada em <i>docking</i> molecular.....	43
1.10.1 Compostos inibidores.....	46
2 Justificativa.....	49
3 OBJETIVOS.....	51
3.1 Objetivo geral.....	51
3.2 Objetivos específicos.....	51
4 METODOLOGIA.....	53
4.1 <b>Metodologia computacional</b> .....	53
4.1.1 Análise in silico da catepsina B de <i>Trypanosoma cruzi</i> .....	53
4.1.2 Modelagem da estrutura tridimensional da catepsina B de <i>Trypanosoma cruzi</i> .....	54
4.1.3 Busca por compostos para validação do protocolo de <i>docking</i> .....	56
4.1.4 Preparo dos ligantes e proteína.....	57
4.1.5 <i>Docking</i> molecular.....	57

4.1.6	Validação do protocolo de <i>docking</i> molecular .....	57
4.1.7	Triagem virtual.....	58
4.1.8	Filtro de eficiência de ligante.....	59
4.1.9	Filtro de QSAR fenotípico .....	60
4.1.10	Filtro de similaridade estrutural e agrupamento e compostos .....	60
4.1.11	Inspeção visual dos compostos .....	60
4.2	<b>Metodologia experimental</b> .....	61
4.2.1	Cultura de células.....	61
4.2.2	Desenho do sgRNA para nocaute com CRISPR/Cas9 .....	61
4.2.3	Amplificação e purificação do sgRNA.....	62
4.2.4	Preparação do plasmídeo pTREX-n/Cas9 e ligação ao sgRNA.....	62
4.2.5	Transformação em bactérias competentes.....	63
4.2.6	Confirmação dos clones transformados de pTREX-n/Cas9/sgRNA.....	63
4.2.7	Produção de cassete para reparo por homologia (HDR).....	64
4.2.8	Amplificação do cassete de reparo por homologia (HDR) .....	65
4.2.9	Clonagem e transformação do cassete amplificado .....	66
4.2.10	PCR de colônia dos transformados com o cassete .....	67
4.2.11	Transfecção do plasmídeo pTREX-n/Cas9/sgRNA em <i>T. cruzi</i> CL-Brener e Y .....	68
4.2.12	Transfecção cassete de blasticidina em <i>T. cruzi</i> nas cepas CL-Brener e Y.....	68
4.2.13	Expressão da catepsina B e análise por <i>Western Blot</i> .....	69
4.2.14	Purificação das frações solúvel e insolúvel da catepsina B .....	71
4.2.15	Concentração das frações solúvel e insolúvel da catepsina B.....	72
5	<b>RESULTADOS</b> .....	74
5.1	Localização dos alelos de catepsina B de <i>T. cruzi</i> na cepa CL Brener .....	74
5.2	Alinhamento das sequências de nucleotídeos dos alelos da catepsina B de <i>T. cruzi</i> e desenho do sgRNA .....	74
5.3	Alinhamento das sequências de aminoácidos dos alelos da catepsina B de <i>T. cruzi</i> .....	77
5.4	Análises <i>in silico</i> da catepsina B de <i>T. cruzi</i> .....	78
5.4.1	Ortologia da catepsina B de <i>T. cruzi</i> no filo Euglenozoa .....	78
5.4.2	Árvore filogenética.....	79
5.4.3	Predição de massa molecular e ponto isoelétrico .....	80
5.4.4	Presença de peptídeo sinal.....	81
5.4.5	Presença de domínio transmembrana.....	82
5.4.6	Motivos de N-glicosilação .....	83
5.4.7	Motivos de O-glicosilação .....	84

5.4.8	STRING.....	85
5.4.9	Alinhamento de sequências de catepsina B de diferentes organismos.....	88
5.4.10	Predição de estrutura secundária da catepsina B de <i>T. cruzi</i> .....	89
5.4.11	Modelagem da estrutura tridimensional da catepsina B de <i>Trypanosoma cruzi</i> .....	90
5.4.12	Validação do protocolo de <i>docking</i> com compostos ativos e <i>decoys</i> .....	94
5.4.13	Triagem virtual dos compostos do Museu Nacional de História Natural de Paris .....	96
<b>PARTE II: DADOS EXPERIMENTAIS .....</b>		<b>102</b>
5.5	Nocautê gênico e amplificação do sgRNA.....	102
5.6	Preparação do plasmídeo pTREX-n/Cas9.....	104
5.7	Amplificação do cassete de homologia .....	105
5.8	PCR de colônia de confirmação do cassete clonado .....	105
5.9	Expressão heteróloga da catepsina B de <i>Trypanosoma cruzi</i> .....	107
5.9.1	Expressão em <i>E. coli</i> Rosetta (DE3) pLysS .....	107
5.9.2	Expressão em <i>E. coli</i> BL21 (DE3) .....	108
5.9.3	Análise das frações solúveis expressas .....	109
5.9.4	Purificação da fração solúvel expressa em <i>E. coli</i> BL21 (DE3).....	110
5.9.5	Concentração da purificação da fração solúvel.....	111
5.9.6	Purificação da fração insolúvel expressa em <i>E. coli</i> Rosetta (DE3) pLysS .....	112
5.9.7	Teste de atividade das porções purificadas .....	113
6	DISCUSSÃO.....	115
7	CONCLUSÃO.....	123
8	PERSPECTIVAS.....	125
REFERÊNCIAS BIBLIOGRÁFICAS.....		127

## LISTA DE ABREVIATURAS

**CRISPR:** *Clustered Regularly Interspaced Short Palindromic Repeats*

**GFP:** Green Fluorescent Protein

**LB:** Meio Luria-Bertani

**MW:** Massa molecular

**Pb:** Pares de base

**PBS:** Phosphate Buffered Saline

**PCR:** Reação em cadeia da polimerase

**RNAi:** RNA interferente

**sgRNA:** RNA guia utilizado no sistema CRISPR/ Cas 9

**SOC:** Meio LB enriquecido com glicose

## LISTA DE FIGURAS

<b>Figura 1. A. Imagem da primeira paciente diagnosticada com a Doença de Chagas, Berenice. B. Notícia do jornal reportando as histórias da doença e da paciente.</b> Adaptado de (LIDANI et al., 2019). .....	19
<b>Figura 2. Número estimado de imigrantes infectados com <i>T. cruzi</i> em países não endêmicos da doença de Chagas baseado em dados dos estados Unidos (2007), Canadá (2006), Europa (2008-2011), Japão (2007), Austrália (2011) e Nova Zelândia (2006).</b> Adaptado de (LIDANI et al., 2019) .....	22
<b>Figura 3. Ciclo biológico do parasito <i>Trypanosoma cruzi</i>.</b> Fonte: CDC Acesso em: 15/01/2020. ....	24
<b>Figura 4. Representação 2D da estrutura química dos medicamentos benznidazol e nifurtimox.</b> ....	25
<b>Figura 5. Modelo da estrutura 3D da catepsina B humana madura (PDB 1HUC) (MUSIL et al., 1991).</b> Figura gerada pelo programa UCSF CHIMERA, destaque aos resíduos do sítio catalítico em vermelho e o <i>loop</i> de oclusão em azul. ....	30
<b>Figura 6. Representação da classificação filogenética das DTUs I-VI de <i>T. cruzi</i> indicando os ancestrais em comum e as formações de cada DTU.</b> Adaptado de (TOMASINI; DIOSQUE, 2015). ....	35
<b>Figura 7. Representação do mecanismo de processamento pós- transcricional do mRNA em <i>T. cruzi</i>.</b> Adaptado de (HERREROS-CABELLO et al., 2020). ....	36
<b>Figura 8. Representação do sistema CRISPR/Cas9 atuando na quebra da dupla fita de DNA dispondo dos componentes moleculares e dos mecanismos de reparo de DNA.</b> Adaptado de (LANDER; CHIURILLO; DOCAMPO, 2016). ....	39
<b>Figura 9. Representação do plasmídeo sgRNA/Cas9/pTREX-n indicando a presença de todos os seus componentes.</b> O sgRNA, na cor rosa, flanqueado pelo promotor ribossomal e o sinal <i>hx1</i> com os sítios de restrição para <i>Bam</i> HI. Além da <i>Sp</i> Cas9 fusionada à GFP e NLS como marca de seleção para neomicina. Adaptado de (DOCAMPO, 2015). ....	40
<b>Figura 10. Representação dos componentes e o resultado do <i>docking</i> com o encaixe de um receptor e um ligante.</b> ....	44
<b>Figura 11. Estrutura 2D de compostos inibidores de catepsina B de <i>S. mansoni</i> (compostos 1-5) <i>T. brucei</i> (compostos 6 e 7), e <i>H. sapiens</i> (compostos 8-14).</b> ....	47
<b>Figura 12. Esquema das etapas da triagem virtual dos compostos provenientes do Museu Nacional de História Natural de Paris (MNHN).</b> As etapas seguem a seguinte sequência: filtro de duplicatas; <i>docking</i> molecular dos compostos contra o modelo 3D da catepsina B de <i>T. cruzi</i> ; filtro de eficiência de ligante (LE); filtro fenotípico de QSAR e inspeção visual.....	59
<b>Figura 13. Esquema do reparo por homologia causado pelos ultrâmeros no local de corte da <i>Sp</i>Cas9.</b> <b>A.</b> Gene <i>catbtc</i> contendo 1002 pb de azul claro com a presença das regiões 5'UTR e 3'UTR do gene de azul escuro. Seta vertical indicando o sítio de corte da <i>Sp</i> Cas9 na posição 331 dentro do gene <i>catbtc</i> . Gene da blasticidina ( <i>bsd</i> ) contendo 399 pb de roxo com a inclusão dos ultrâmeros provenientes da região 5'UTR <i>catbtc</i> (azul escuro) e após o sítio 331 dentro do gene <i>catbtc</i> (azul claro) indicando a região de reparo por homologia direta realizada após o corte da <i>Sp</i> Cas9. <b>B.</b> Gene da blasticidina incorporado dentro do gene da <i>catbtc</i> contendo 1052 bp. ....	65
<b>Figura 14. Representação do vetor pGEM®-T Easy (Promega™) utilizado na clonagem do cassete visto no programa SnapGene importado das sequências online do próprio programa.</b> Representação dos componentes desse vetor incluindo o sítio de múltipla clonagem. ....	67

- Figura 15. Representação do plasmídeo pET-19b contendo suas regiões destacadas feito no SnapGene importado das sequências depositadas online do próprio programa.** Na imagem está incluso o MCS, sítio de múltipla clonagem, onde a clonagem da pró-catepsina B foi realizada.....71
- Figura 16. Alinhamento dos alelos *Esmeraldo Like* e *Non-Esmeraldo Like*, respectivamente, da catepsina B de *Trypanosoma cruzi* na cepa CL Brener feito no MUSCLE.** Os asteriscos indicam identidade entre os nucleotídeos de ambos os alelos e sua ausência, indica falta de identidade. A região vermelha destacada em vermelho indica a sequência de 20 nucleotídeos nas posições 331-350 escolhida para produzir o RNA guia.....76
- Figura 17. Alinhamento dos aminoácidos correspondentes aos *Esmeraldo Like* e *Non-Esmeraldo Like*, respectivamente, da catepsina B de *Trypanosoma cruzi* na cepa CL Brener feito no MUSCLE.** Região do peptídeo sinal destacada de azul entre os resíduos 1-21, região da pró- catepsina B de cinza entre os resíduos 22-91, domínio Peptidase\_C1 da proteína entre os resíduos 92-329 destacado de amarelo, a tríade catalítica representada de verde pelos resíduos de Cisteína 120, Histidina 278 e Asparagina 298, o *loop* de oclusão que se estende entre os resíduos 187-209 representado de rosa. Os aminoácidos que possuem o asterisco obtiveram identidade entre si, os representados por 2 pontos possuem fortes propriedades bioquímicas similares, os que possuem 1 ponto compartilham fracas propriedades bioquímicas similares e os que não possuem marcação não compartilham propriedade bioquímica similar.....78
- Figura 18. Descrição da ortologia do gene codificador da proteína catepsina B de *T. cruzi* da cepa CL Brener no filo Euglenozoa feito no OrthoDB.** .....79
- Figura 19. Árvore filogenética produzida a partir de 24 proteínas obtidas no BLAST com a catepsina B de *T. cruzi* gerada no IQ-Tree Web Server e visualizada no FigTree 1.4.4.** São indicados os valores de *likelihood* e *bootstrap* gerados em cada nó. O nodo referente à catepsina B de *T. cruzi* cepa CL Brener está destacado de vermelho. A linha na base da imagem indica o número total de mutações dividido pelo total de bases do alinhamento das sequências.....80
- Figura 20. Predição do número de aminoácidos, massa molecular e ponto isoelétrico teórico da *catbtc* completa codificada pelo alelo *Esmeraldo-like* feito no ProtParam.....**81
- Figura 21. Predição do sítio de clivagem de peptídeo sinal da *catbtc* completa codificada pelo alelo *Esmeraldo-like* feito no SignalP 5.0.** .....82
- Figura 22. Predição do domínio transmembrana da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no TMHMM 2.0.** .....83
- Figura 23. Predição de motivos de N-glicosilação da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no NetNGlyc 1.0.** Motivos potenciais (azul) e aminoácidos preditos (vermelho) como sítios de N-glicosilação,.....84
- Figura 24. Predição de motivos de O-glicosilação da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no NetOGlyc 4.0.**.....85
- Figura 25. Interações realizadas pela catepsina B descritas com proteínas relacionadas ao *T. cruzi* feito no STRING.** As cores indicam o *score* de interação com as outras proteínas variando em um mapa contínuo de amarelo a roxo. As linhas mais grossas apontam interações com maior grau de confiança. ....87
- Figura 26. Interações realizadas pela catepsina B descritas com proteínas relacionadas a *Homo sapiens* feito no STRING.** As cores indicam a presença de interação com proteínas descritas no lisossomo, variando em um mapa contínuo de amarelo a roxo. As linhas mais grossas apontam interações com maior grau de confiança. ....88

- Figura 27. Alinhamento das sequências de catepsina B de *H. sapiens* (CTSB), *T. brucei* (CatBTb) e *T. cruzi* (CatBTc).** As cores vermelhas indicam resíduos que possuem 100% de identidade nas três sequências, as cores amarelas, fortes propriedades bioquímicas similares e os números de verde, as pontes de dissulfeto com seus respectivos pares. O sítio ativo está presente nos resíduos 108, 278 e 298. ....89
- Figura 28. Predição da estrutura secundária da catepsina B de *T. cruzi* a partir da sequência primária no servidor NetSurfP-2.0.** Na imagem, as voltas laranjas indicam regiões de predição de  $\alpha$ -hélice; as setas roxas, folhas  $\beta$ ; as linhas vermelhas, exposição à superfície; as linhas azuis, não-exposição à superfície e linhas cinzas, o nível de desordem dos resíduos pela espessura da linha. ....90
- Figura 29. Modelos tridimensionais da catepsina B de *Trypanosoma cruzi*.** **A.** Pró-catepsina B, gerada pelo servidor trRosetta. **B.** Catepsina B madura, gerada pelo servidor trRosetta. **C.** Catepsina B completa, gerada pelo servidor AlphaFold. **D.** Catepsina B madura, gerada pelo servidor AlphaFold. **E.** Sobreposição das catepsina B maduras, geradas pelos servidores trRosetta (bege) e AlphaFold (azul). Os resíduos do sítio catalítico e *loop* de oclusão (representados em *sticks*) dos modelos trRosetta e AlphaFold foram destacados em vermelho e azul, respectivamente. No centro da imagem se encontram os resíduos do sítio catalítico (Cys120, His278 e Asn298) e na parte superior, os do *loop* de oclusão, com destaque para os resíduos de histidina (His192 e His193) importantes para manutenção da interação com ligantes. Figuras geradas no programa UCSF CHIMERA. ....94
- Figura 30. Gráfico violino da distribuição dos scores de docking dos compostos ativos (rosa) e decoys (verde),** com valor exato e significativo de  $P = < 0,0001$ . Foi aplicado o teste não paramétrico *Mann-Whitney*, apresentando mediana dos ativos em -5,962 e  $n = 15$ ; enquanto a mediana dos inativos foi de -4,584 e  $n = 521$ . A linha preta central mais grossa representa a mediana e as linhas pretas mais finas superiores e inferiores representam o 3º quartil e 1º quartil, respectivamente. ....96
- Figura 31. Esquema do fluxograma da triagem virtual dos compostos provenientes do Museu Nacional de História Natural de Paris contra a CatB *T. cruzi*, utilizando diferentes abordagens computacionais.** ....99
- Figura 32. Esquema do diagrama 2D das interações dos ligantes co-cristalizados dos PDBs 1GMV (CatB *H. sapiens*), 3A18 (CatB *H. sapiens*), 3HHI (CatB *T. brucei*) e 6AY2 (CatB *H. sapiens*), utilizados como comparação na inspeção visual dos resultados de docking da catepsina B *T. cruzi*.** ....100
- Figura 33. Esquema do diagrama 2D das interações dos top 10 compostos e fragmentos selecionados com a CatBTc, apresentando os 6 primeiros.** Figuras geradas pelo servidor ProteinsPlus, com a ferramenta PoseView e compostos mostrados de acordo com o score de docking em ordem crescente. Continua na Figura 32. ....101
- Figura 34. Conclusão do esquema do diagrama 2D das interações dos top 10 compostos e fragmentos selecionados com a CatBTc, apresentando os 4 últimos.** Figuras geradas pelo servidor ProteinsPlus, com a ferramenta PoseView e compostos mostrados de acordo com o score de docking em ordem crescente. ....102
- Figura 35. A. Eletroforese em gel de agarose 2,5% da amplificação do sgRNA.** Marcador de peso molecular 1 Kb plus Invitrogen™ utilizado no poço 1; sgRNA no poço 3 com 122 pb. **B. Eletroforese em gel de agarose 1% da linearização do plasmídeo pTREX-n/Cas9.** Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; poço 2, o plasmídeo antes da digestão e poço 3 com o plasmídeo digerido com *Bam*HI, ambos com 11,2k pb. ....103
- Figura 36. Eletroforese em gel de agarose 2,5% da confirmação da orientação correta do sgRNA clonado no pTREX-n/sgRNA/Cas9.** Os poços foram nomeados com letras ordenadas alfabeticamente de A-M. Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço A; as

- colônias estão distribuídas entre os poços B-L e o controle negativo está no poço M. sgRNA amplificado com tamanho de 190 pb nos poços I e K. ....104
- Figura 37. Eletroforese em gel de agarose 1,5% da amplificação do cassete de homologia.** Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; poços 2 e 3 contendo diferentes concentrações do cassete amplificado com 556 pb. O poço 2 mostra a concentração correspondente a 2µl de cassete amplificado, enquanto o poço 3 corresponde a 1µl.....105
- Figura 38. Eletroforese em gel de agarose 2,5% da confirmação do cassete clonado no pGEM®-T Easy.** Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; O cassete amplificado possui 181 pb. Os poços 2-5 representam colônias brancas do cassete clonado no pGEM®-T Easy; os poços 6-9 são colônias brancas do pGEM-BSD; o poço 10 é uma colônia azul do cassete clonado no pGEM®-T Easy; o poço 11 não possui DNA, apenas os *primers*. ....106
- Figura 39. Indução da pró-catepsina B com IPTG 0,5mM à 20°C, por 5 horas em *E. coli* Rosetta (DE3) pLysS. A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução da pró-catepsina B com o poço 1 com a fração solúvel, o poço 2 com a fração insolúvel e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B. Western Blot** da indução da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. O poço 1 representa a fração solúvel, o poço 2 a fração insolúvel e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. ....108
- Figura 40. Indução da pró-catepsina B com IPTG 1mM à 37°C, por 3 horas em *E. coli* BL21 (DE3). A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução da pró-catepsina B com os poços 1-3 com frações solúveis, os poços 4-6 com frações insolúveis e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B. Western Blot** da indução da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. Os poços 1-3 representam as frações solúveis, os poços 4-6, as frações insolúveis e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. ....109
- Figura 41. Indução das frações solúveis previamente expressas da pró-catepsina B em *E. coli* cepas BL21 (DE3) e Rosetta (DE3) pLysS. A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução das frações solúveis da pró-catepsina B com os poços 1-2 correspondentes às expressões em BL21 (DE3); poço 3, em Rosetta (DE3) pLysS e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B. Western Blot** da indução das frações solúveis da pró-catepsina B com os poços 1-2 correspondentes às expressões em BL21 (DE3); poço 3, em Rosetta (DE3) pLysS e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. ....110
- Figura 42. Purificação da fração solúvel da pró-catepsina B a partir da expressão em *E. coli* BL21 (DE3). A.** Gel SDS-PAGE 12% corado com azul de Coomassie da purificação da pró-catepsina B com o poço 1 contendo o extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poços 5-9, eluições com imidazol 10mM, 20mM, 40mM, 60mM e 80mM, respectivamente; poços 10-14, eluições com imidazol 250mM e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B. Western Blot** da purificação da fração solúvel da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 contendo o extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço4, tampão de ligação final; poços 5-9, eluições com imidazol 10mM, 20mM, 40mM, 60mM e 80mM, respectivamente; poços 10-14, eluições com imidazol 250mM e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. ....111
- Figura 43. Concentração da purificação da fração solúvel da pró-catepsina B a partir da expressão em *E. coli* BL21 (DE3). A.** Gel SDS-PAGE 12% corado com azul de Coomassie da concentração da

purificação da pró-catepsina B com o poço 1 contendo 80-; o poço 2, 80+; poço 3, L80-; poço4, L80 e MW, marcador de massa molecular PageRuler™ Plus Prestained Protein Ladder, Thermo Scientific™.

**B. Western Blot** da purificação da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 contendo 80-; o poço 2, 80+; poço 3, L80-; poço4, L80 e MW, marcador de massa molecular PageRuler™ Plus Prestained Protein Ladder, Thermo Scientific™. ....112

**Figura 44. Purificação da fração insolúvel da pró-catepsina B a partir da expressão em *E. coli* Rosetta (DE3) pLysS.**

**A.** Gel SDS-PAGE 12% corado com azul de Coomassie da purificação da pró-catepsina B com o poço 1 equivalente ao extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poço 5, eluição com imidazol 10mM; poços 6-12, eluições com imidazol 250mM e MW, marcador de massa molecular BenchMark™ Protein Ladder, Thermo Scientific™.

**B. Western Blot** da purificação da fração insolúvel da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 equivalente ao extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poço 5, eluição com imidazol 10mM; poços 6-12, eluições com imidazol 250mM e MW, marcador de massa molecular BenchMark™ Protein Ladder, Thermo Scientific™. ....113

## LISTA DE TABELAS

<b>Tabela 1</b> - <i>Primers</i> utilizados na metodologia CRISPR/Cas9.....	62
<b>Tabela 2</b> - <i>Primers</i> usados na metodologia de amplificação e confirmação do cassete de homologia	65
<b>Tabela 3</b> - Identificação e localização dos alelos do gene <i>catbtc</i> .....	74
<b>Tabela 4</b> - Modelos 3D da CatB de <i>T.cruzi</i> .....	92
<b>Tabela 5</b> - Métricas de validação dos cálculos de <i>docking</i> dos compostos ativos e <i>decoys</i> nos modelos CatBTC (trRosetta e AlphaFold), realizados nos servidores DockThor e Glide.....	95
<b>Tabela 6</b> - Compostos priorizados e seus respectivos <i>scores</i> de <i>docking</i> , eficiência do ligante (LE) e probabilidade de atividade frente a <i>T. cruzi</i> (predita pelo modelo QSAR fenotípico).....	97

## RESUMO

A doença de Chagas tem como agente etiológico o parasito intracelular obrigatório *Trypanosoma cruzi* e acomete aproximadamente 7 milhões de pessoas no mundo sendo endêmica em países da América Latina. A doença caracteriza-se ainda como negligenciada por não possuir tratamento eficaz para os casos durante a fase crônica, bem como cura e por afetar, em sua maioria, a população de baixa renda. Em vista disso, é necessária a descoberta de fármacos mais eficazes e seguros com novos mecanismos de ação. Nesse contexto, a catepsina B destaca-se como uma cisteíno protease essencial em outros organismos pela sua atuação no processo infeccioso. Embora já tenha sua atividade enzimática estudada, abordagens de estrutura e função da catepsina B de *T. cruzi* (CatBTc) não haviam ainda sido exploradas. Este trabalho, portanto, teve como objetivo a elucidação das funções da catepsina B de *T. cruzi* por meio de análises *in silico* e *in vitro*. Dessa forma, desenvolvemos e avaliamos modelos tridimensionais da CatBTc. A partir do melhor modelo obtido, realizamos uma triagem virtual da biblioteca de compostos do Museu Nacional de História Natural de Paris, por meio de cálculos de *docking* molecular e outras abordagens computacionais para priorizar compostos potenciais inibidores de CatBTc para posterior validação *in vitro*. Por fim, a obtenção da proteína recombinante expressa em *E. coli* e do nocaute duplo dos alelos presentes nas cepas CL Brener e Y utilizando o sistema CRISPR/Cas9 foram explorados com a finalidade de realizar a sua análise funcional. Assim, foi possível se obter dois modelos tridimensionais computacionais da protease provenientes de servidores diferentes que foram submetidos a cálculos de *docking* com a seleção 19 compostos promissores. A proteína recombinante foi expressa e purificada nas frações solúvel e insolúvel com sucesso, contudo, não apresentou atividade. A primeira tentativa de nocaute gênico resultou na morte e dificuldade de seleção de parasitos mutantes possivelmente devido importância do gene para a sobrevivência do parasito.

**Palavras-chave:** doença de Chagas; proteases; catepsina B; *docking* molecular; inibidores; modelagem estrutural; triagem virtual; expressão de proteína heteróloga; CRISPR/Cas9

## ABSTRACT

Chagas Disease is caused by the obligate intracellular parasite *Trypanosoma cruzi* and affects approximately 7 million people worldwide. It is considered a neglected tropical disease (NTD) for mostly infect low-income population from Latin American countries and not having effective and safe treatment for the most severe cases. Therefore, new therapeutic alternatives are required. In this sense, proteases stand out as potential molecular targets of drugs for acting in essential biological processes. Among them, the cysteine protease, cathepsin B sticks out in other organisms for its role in infectious process. Although its enzymatic activity has already been studied, structural and functional approaches of cathepsin B *T. cruzi* have not been explored yet. Hence, this work, aimed to enlighten the performance and functions of cathepsin B of *T. cruzi* through in silico and in vitro analyses. In this way, we developed and evaluated three-dimensional CatbTc models. From the best obtained 3D model, we performed a virtual screening of the compound library of the *Muséum National d'Histoire Naturelle de Paris* (MNHN), through molecular *docking* calculations and other computational approaches to prioritize potential CatbTc inhibitor compounds for further in vitro validation. Ultimately, we attempted to obtain the recombinant protein expressed in *E. coli* and the double knockout of the alleles present in the strains CL Brener and Y using the CRISPR/Cas9 system aspiring to perform its functional analysis. Thus, it was possible to obtain two 3D models of the protease from distinct servers that were submitted to *docking* calculations, finally selecting 19 promising compounds. The recombinant protein was successfully expressed and purified in soluble and insoluble fractions, however, it displayed no activity. The first gene knockout attempt resulted in the death and difficulty of selecting mutant parasites, which was likely due to the relevance of the gene for the parasite's survival.

**Keywords:** Chagas disease; proteases; cathepsin B; molecular *docking*; inhibitors; structural modelling; heterologous protein expression; CRISPR/Cas9

# INTRODUÇÃO

# 1 INTRODUÇÃO

## 1.1 Doença de Chagas

A doença de Chagas (DC) foi descoberta em 1909 pelo pesquisador brasileiro Carlos Chagas, que foi capaz de descrever a enfermidade completa contendo suas manifestações clínicas, epidemiológicas, o parasito e o seu vetor em sua paciente criança chamada Berenice (**Figura 1**) (CHAGAS, 1909). A doença é uma antropozoonose causada pelo protozoário flagelado *Trypanosoma cruzi* que possui em seu ciclo biológico, hospedeiros vertebrados e invertebrados (NEVES, 2004).



**Figura 1. A. Imagem da primeira paciente diagnosticada com a Doença de Chagas, Berenice. B. Notícia do jornal reportando as histórias da doença e da paciente.** Adaptado de (LIDANI et al., 2019).

No que concerne o contágio da DC, a transmissão vetorial é descrita classicamente desde os primórdios da descoberta da doença, porém outras formas ocorrem por outras vias como: transfusão sanguínea de indivíduos infectados, transmissão vertical durante a gestação e por intermédio oral (WHO, 2021). Atualmente, a transmissão por via oral envolvendo alimentos, como cana-de-açúcar e açaí, representa a principal via de infecção humana. Há registros de surtos atribuídos a contaminação oral em todas as áreas rurais e periurbanas da América Latina (SHIKANAI-YASUDA; CARVALHO, 2012). As formas secundárias se enquadram em

acidentes de laboratório, durante a manipulação de amostras ou animais infectados e doação de órgãos (COURA, 2015).

A DC é dividida classicamente em duas fases distintas no hospedeiro vertebrado, apresentando assim, fase aguda e fase crônica, onde a primeira apresenta quadro assintomático ou inespecífico com a presença de dor de cabeça, dores no corpo, febre e dificuldade para respirar e pode ser facilmente diagnosticada se testada, pois ocorre a presença de grande quantidade do parasita no sangue periférico (DE SOUZA et al., 2016; STANAWAY; ROTH, 2015). Durante essa fase, alguns pacientes podem apresentar o chagoma, ou seja, um edema ou sinal de Romaña, um clássico nódulo inflamatório de inchaço periorbital quando o *T. cruzi* tem acesso pela conjuntiva (MALIK; SINGH; AMSTERDAM, 2015). Já a fase crônica tem início cerca de 2 a 4 meses após a infecção aguda, concomitantemente com a diminuição drástica da presença do parasito na corrente sanguínea. Inicia-se então um período silencioso prolongado, denominado forma indeterminada, no qual não se encontram manifestações clínicas significantes além da sorologia positiva (MACIEL, 2016; (STANAWAY; ROTH, 2015). Tal configuração pode persistir por toda a vida pois, aproximadamente 20-40% dos casos evoluem para as manifestações clínicas predominantes nos sistemas cardiovascular e digestório (DUMONTEIL; HERRERA, 2017). Dentre essas manifestações, ocorrem taquicardias, bradicardias, aumento do órgão que levam à parada cardíaca. Ainda, os sintomas podem evoluir também para o desenvolvimento de megacólon e/ou megaesôfago, em que esses órgãos expandem radicalmente de tamanho. Entretanto, os sintomas cardíacos são os mais presentes e severos, acometendo cerca de 1/3 dos indivíduos com sorologia positiva, dado que o parasito possui maior tropismo para esse tipo de tecido. Os casos mais graves de cardiomiopatia não possuem algum tratamento senão o transplante do órgão (BERN, 2011; LIDANI et al., 2019).

## 1.2 Distribuição da doença

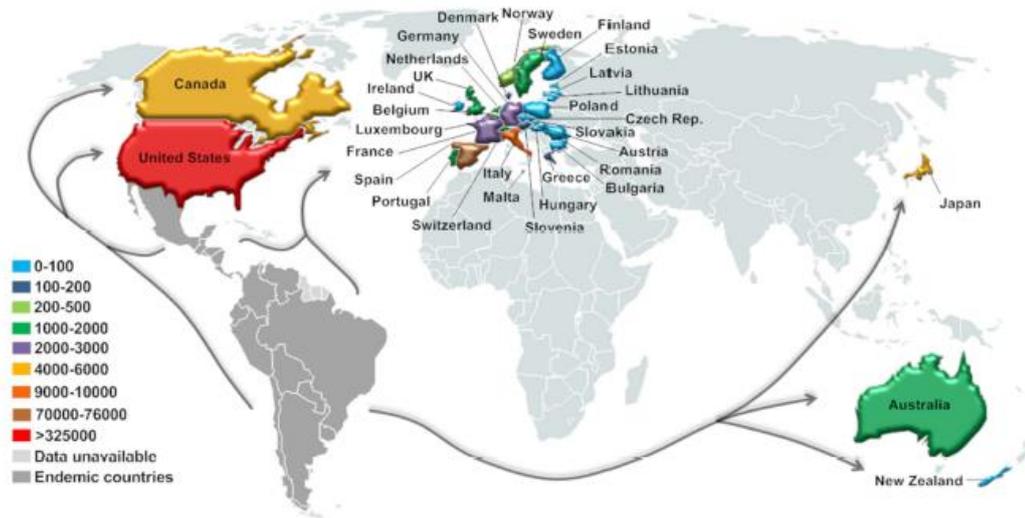
A DC é atualmente considerada pela Organização Mundial da Saúde como uma doença tropical negligenciada, pois acomete, em sua maioria, populações de baixa renda, com estimativa de 7 milhões de pessoas infectadas em todo o mundo. É considerada também uma enfermidade endêmica de regiões tropicais e é encontrada

principalmente em 21 países da América latina (WHO, 2021). A DC ainda é considerada um grande problema de saúde pública tanto no Brasil como na América Latina, segundo o Ministério da Saúde, no período de 2000 a 2013 foram registrados de 1.570 casos de doença de Chagas aguda no Brasil, com a região norte detendo 91,1% dos casos, sendo a transmissão oral como a via mais frequente, aproximadamente 69% dos casos totais (BRASIL, 2015). De acordo com a Organização Mundial da Saúde (OMS), em 2018, o número estimado de pessoas com DC, aguda ou crônica, no Brasil ultrapassa 900 000, sendo o líder de casos junto com Argentina e México na América Latina (WHO, 2018).

A forma oral tem sua maior concentração na região amazônica, onde ocorreram surtos de infecção, correspondentes à ingestão de alimentos contaminados com as fezes do vetor e/ou com partes do triatomíneo e que não foram devidamente higienizados como por exemplo: açaí, caldo de cana e sucos fazem parte dos grandes contribuidores dessa transmissão (SHIKANAI-YASUDA; CARVALHO, 2012). O impacto social, político e econômico da doença é alto, pois é estimado que sejam gastos 1,2 bilhão de dólares anualmente na América Latina devido à perda de produtividade, sendo considerada, na América Latina, a doença parasitária que mais apresenta essa repercussão (VIEIRA; PRAÇA; LUÍS, 2018).

Atualmente, os movimentos migratórios acabaram por causar uma redistribuição da doença em países como: Estados Unidos, Japão, Canadá, Espanha, França, Suíça, Nova Zelândia e Austrália, por exemplo, como é possível verificar na **Figura 2**. Nos Estados Unidos, por exemplo, até 2015 só haviam sido reportados 28 casos de transmissão vetorial desde o primeiro relato de um caso em 1955. Porém, esse número não representa o total real de pessoas infectados no país, visto que a maioria dos casos reportados são por meio de imigrantes que acessam o país já portando a doença (LIDANI et al., 2019). Estudos recentes de vigilância mostraram que triatomíneos adultos e ninfas foram encontrados em regiões domésticas e peri domésticas no sul dos Estados Unidos, elevando o número dos casos de transmissão vetorial (VIEIRA; PRAÇA; LUÍS, 2018). Outros fatores relacionados aos aumentos de casos são causados também pela exportação de alimentos, transplante de órgãos, transfusão sanguínea e transmissão congênita do parasito. Estes dados constataam que em países considerados não endêmicos, existe uma grande falha nas políticas de controle do vetor (JACKSON, 2016). A razão para esses procedimentos passarem despercebidos é explicada pela baixa ocorrência e, portanto, falha na triagem e

fiscalização nesses países. Assim, constitui-se uma população sujeita às transmissões verticais e congênitas (MANNE-GOEHLER et al., 2016).



**Figura 2.** Número estimado de imigrantes infectados com *T. cruzi* em países não endêmicos da doença de Chagas baseado em dados dos estados Unidos (2007), Canadá (2006), Europa (2008-2011), Japão (2007), Austrália (2011) e Nova Zelândia (2006). Adaptado de (LIDANI et al., 2019)

### 1.3 Ciclo biológico do *Trypanosoma cruzi*

O parasito *Trypanosoma cruzi* pertence à ordem Kinetoplastida, família Trypanosomatidae sendo, portanto, um eucarioto hemoflagelado. Está inserido no gênero *Trypanosoma* do qual existem diversos exemplos de organismos, os quais são capazes de infectar e causar doenças em humanos e outros mamíferos (BRENER, 1997). *T. cruzi* se caracteriza como um parasito intracelular obrigatório e, dentro do seu ciclo biológico, o agente etiológico possui 4 formas distintas que apresentam características bioquímicas, morfológicas e metabólicas particulares, as quais são expressas diferencialmente de acordo com o hospedeiro infectado, já que o protozoário alterna o ciclo entre um hospedeiro invertebrado e um vertebrado (BRINGAUD; RIVIÈRE; COUSTOU, 2006).

O principal vetor do *T. cruzi* é o seu hospedeiro invertebrado, um inseto chamado triatomíneo, ou popularmente barbeiro no Brasil, pertencente à família *Reduviidae*, subfamília Triatominae. Ele se alimenta do sangue do hospedeiro vertebrado, pois apresenta comportamento hematófago para repor necessidades nutricionais necessárias para seu devido crescimento (SILVA, 2009). Todas as espécies desta subfamília estão aptas para a transmissão do parasito ao hospedeiro

definitivo, mas os relativos aos gêneros *Triatoma*, *Rhodnius* and *Panstrongylus* tem maior importância epidemiológica na América Latina, em razão da adaptação em colonizar ambientes peridomiciliar e domiciliar, entrando em maior contato com animais domésticos e humanos (VIEIRA; PRAÇA; LUÍS, 2018).

As formas do parasito se desenvolvem alternando entre estágios não infectivos e infectivos. As formas replicativas, conhecidas como epimastigostas e amastigostas, estão localizadas no intestino do vetor triatomíneo e na célula do hospedeiro mamífero, respectivamente. Em contrapartida, as formas infectivas, chamadas tripomastigostas sanguíneos e tripomastigostas metacíclicos são encontradas, respectivamente, na corrente sanguínea de mamíferos e na porção final do intestino de triatomíneos (SOLÍS-OVIEDO et al., 2018).

O ciclo biológico se inicia com o repasto sanguíneo do triatomíneo infectado com *T. cruzi* em um hospedeiro humano saudável, como pode ser observado na **Figura 3**, assim o inseto tem potencial de transmitir as formas tripomastigotas metacíclicas do parasito, que possuem alta infectividade para o hospedeiro vertebrado. Enquanto ocorre o repasto sanguíneo, o triatomíneo defeca e, assim, consegue disseminar os parasitos através das fezes que podem entrar em contato com o local da picada ou alguma mucosa próxima, acessando a corrente sanguínea. Então, os tripomastigotas metacíclicos invadem as células e se replicam na forma de amastigostas até se modificarem novamente para a forma tripomastigota sanguínea. Nesta etapa, esses parasitos ficam circulantes na corrente sanguínea assumindo a função de infectar novas células e assim criar um ciclo infectivo (C. ONYEKWELU, 2019). O contato do triatomíneo com as formas circulantes do parasito acontece quando ocorre o repasto sanguíneo em um hospedeiro infectado. Assim, os parasitos migram para o intestino do triatomíneo, já transformados em epimastigostas, os quais se multiplicam e se fixam em membranas presentes na porção final do órgão. Este passo é essencial para a maturação deles, chamada metaciclogênese, possibilitando a transformação para a forma infectiva. Esta nova forma expressa um conjunto de glicoproteínas de superfície implicadas em invasão celular em mamíferos (SOLÍS-OVIEDO et al., 2018).

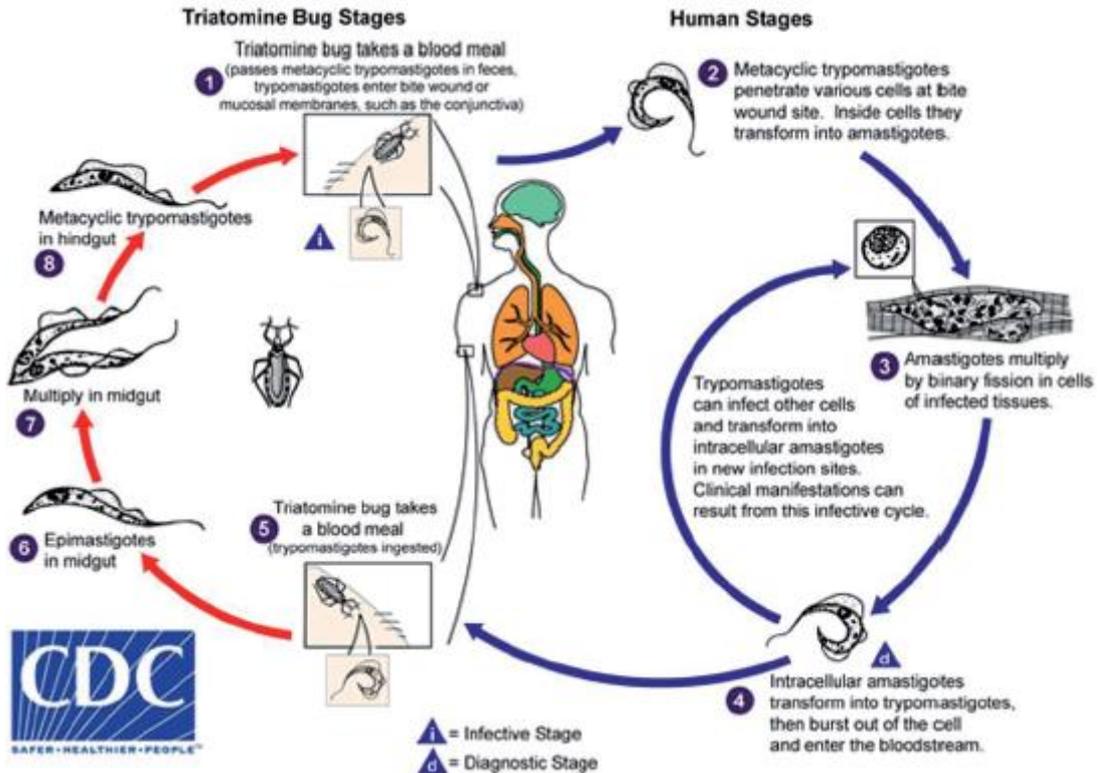


Figura 3. Ciclo biológico do parasito *Trypanosoma cruzi*. Fonte: CDC Acesso em: 15/01/2020.

#### 1.4 Tratamento da doença de Chagas

O tratamento, atualmente, pode ser feito apenas com dois medicamentos, benznidazol e nifurtimox (**Figura 4**), porém apenas o benznidazol é permitido no Brasil. Ambos os fármacos possuem alta capacidade de cura da doença, 60 a 80% dos casos, se administrados nas fases iniciais, ou seja, na fase aguda, apresentando diminuição na eficácia com a progressão da doença, se tornando ineficazes no tratamento durante a fase crônica (MALIK; SINGH; AMSTERDAM, 2015). Foi capaz de se provar que benznidazol não tem potencial de cura em pacientes com cardiomiopatia estabelecida (JR; MARIN-NETO; RASSI, 2017). Um dos maiores problemas envolvidos com esses medicamentos é que eles demonstraram diversos efeitos colaterais tóxicos para o organismo (DUMONTEIL; HERRERA, 2017). O nifurtimox produz nitroreduções, ciclos redox e gera espécies reativas de oxigênio (ROS) que são compostos nocivos para o hospedeiro. Ainda, a espécie reduzida do fármaco é conhecidamente carcinogênica e mutagênica durante o uso prolongado. O benznidazol é metabolizado em espécies reativas parecidas com os formados pelo nifurtimox, assim como é igualmente considerado carcinogênico (BUSCHINI, 2009).

Além disso, o *T. cruzi* possui mecanismo de metabolismo de nitroheterocíclicos pelo gene NTR (nitroredutase), principais compostos do nifurtimox e benznidazol. Dentro do seu mecanismo de ação, os nitroheterocíclicos precisam desse metabolismo por parte do parasito para serem ativos, porém o efeito de resistência a eles já foi descrito com a redução de expressão do gene NTR. Portanto, a possibilidade de uma mutação em que fossem produzidos parasitos com resistência aos únicos medicamentos existentes é preocupante e não pode ser ignorada (WILKINSON et al., 2008).

A melhor forma de prevenção da doença na América Latina é através do controle vetorial evitando um crescimento do contágio em regiões que são típicas, como também a triagem sanguínea de pacientes propensos a transfusões ou transplantes (WHO, 2021). Então, é possível afirmar a necessidade da busca por novos protótipos tripanocidas com mecanismo de ação novo para o tratamento da DC (BUSCHINI, 2009).

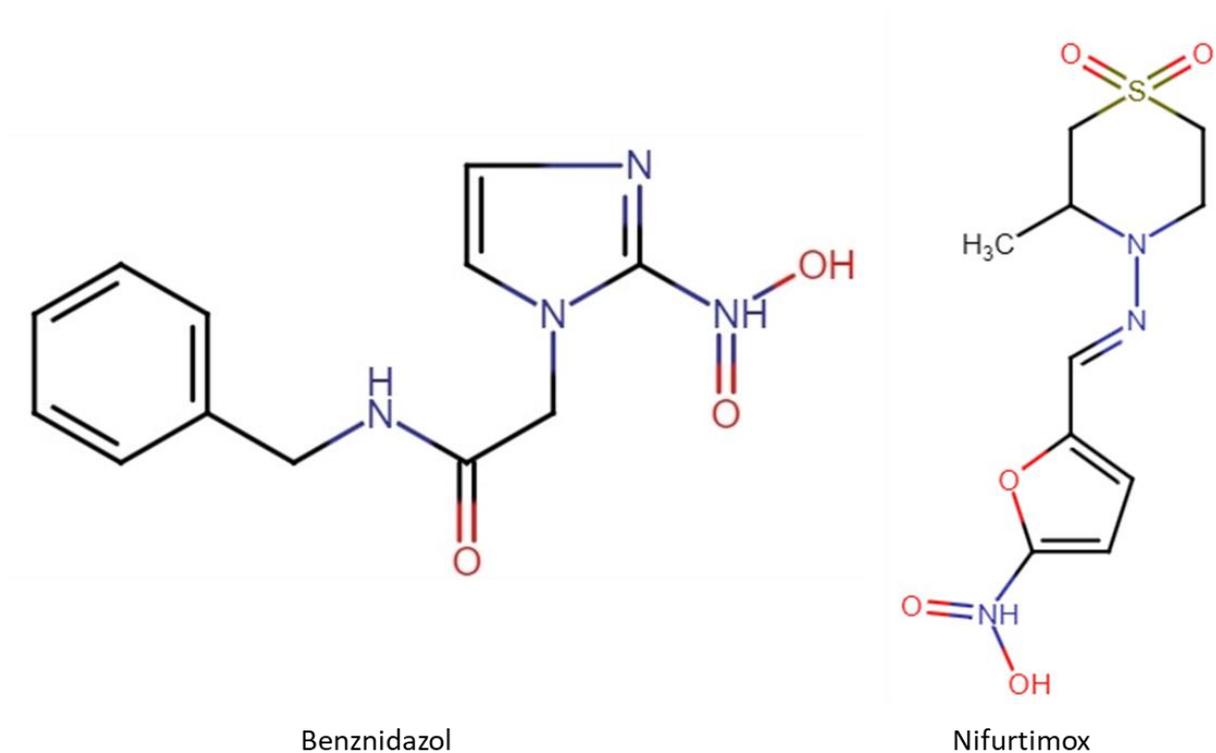


Figura 4. Representação 2D da estrutura química dos medicamentos benznidazol e nifurtimox.

## 1.5 Proteases

Proteases ou peptidases são moléculas que promovem a clivagem através da hidrólise de ligações peptídicas presentes em proteínas e polipeptídeos, transformando-os em resíduos de aminoácidos ou polipeptídeos menores. Elas são enzimas encontradas em diversos organismos e possuem múltiplas funções, dentre elas: migração celular, apoptose, remodelamento e morfogênese tecidual. Partindo desse princípio, são classificadas de acordo com o grupo nucleofílico participante do seu sítio ativo. As proteases são conhecidas atualmente pela presença em pelo menos 7 famílias em que são classificadas como: serino, metalo, cisteíno, aspartato, glutâmico, treonino proteases e asparagina liase (SIQUEIRA-NETO et al., 2018). As peptidases apresentam duas atividades distintas: exopeptidase e endopeptidase. As exopeptidases tem a capacidade de catalisar a formação de aminoácidos livres ou peptídeos, sejam eles na região N-terminal ou C-terminal do substrato, sendo chamadas de aminopeptidases e carboxipeptidases, respectivamente. Enquanto que as endopeptidases hidrolisam ligações peptídicas específicas dentro da cadeia polipeptídica (RAKSAKULTHAI; HAARD, 2003). Elas são capazes de clivar ligações peptídicas de proteínas causando modificações em seus substratos e, conseqüentemente, mudanças biológicas importantes. Entretanto, a atuação das proteases não abrange apenas a degradação de proteínas, mas também o papel de regular a localização e atividade de diversas outras proteínas. Ainda, estão envolvidas na viabilidade de inúmeras células, operando na replicação, transcrição, proliferação, diferenciação e remodelagem da matriz extracelular, por exemplo (BOND, 2019). Uma grande característica delas é sua versatilidade e devido a esta atividade, possuem também atuação considerável em processos patológicos, sendo assim relatadas em quadros de inflamação, doenças auto imunes e neurodegenerativas. Igualmente são descritas em processos de tumorigênese, visto que por estarem relacionadas a diversos processos biológicos representam também um problema quando o equilíbrio é rompido (KAPPELHOFF et al., 2017).

Em humanos, elas são expressas em diferentes tipos celulares, porém foi observada a preferência por alguns tipos específicos como: micróglia, eritrócitos, linfócitos, macrófagos, células dendríticas, epitélio pulmonar e do trato gastrointestinal, por exemplo. Existe também uma relação direta do aumento da

expressão de enzimas particulares nestes tecidos com a ocorrência do desequilíbrio de proteases (PATEL et al., 2018). A localização das proteases ajuda na indicação da sua possível função, sendo assim, expressões diferenciais dessas enzimas são utilizadas como biomarcadores para diagnósticos de certos tipos de patologia, sendo possível relacioná-las como alvos terapêuticos (YADATI et al., 2020). Quando observados em organismos infectantes, apresentam importante papel nas regulações e sinalizações celulares, tanto normais quanto patológicas do parasito, podendo assim ser críticas no processo de infecção do patógeno (TURK et al., 2012; ZHANG et al., 2013).

## 1.6 Catepsinas

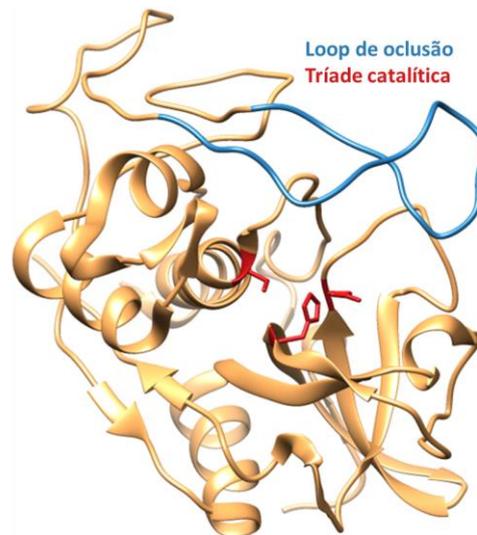
Uma das famílias de proteases que se destacam são as cisteíno proteases. Estas peptidases caracterizam-se por possuírem um resíduo de cisteína (Cys), onde o grupo tiol deste aminoácido atua como nucleofílico na catálise. Geralmente, o sítio catalítico desta enzima contém também um resíduo de histidina (His) que funciona como uma base (BARRETT; RAWLINGS, 1996). Elas são um dos mais famosos e explorados grupos dentre as famílias de proteases, visto que seus mecanismos de ação, assim como de inibição já são bastante elucidados em diversos organismos, sendo envolvidos em diferentes processos patológicos aos seres humanos e representam, portanto, potenciais alvos de fármacos, o que os tornam muito atrativos (MCKERROW, 2018). Os principais exemplos de cisteíno proteases são as catepsinas, que consistem em proteases lisossômicas e, logo, o pH ótimo de atuação é baixo. Por ter encargo de degradação proteica, é necessário um controle estrito da sua atividade. Para tanto, a maioria delas se apresenta inativada em forma de zimogênio demandando uma ativação que faz com que sua conformação seja modificada. Este processo é possível pela presença de formas pré e/ou pró catepsinas, que são posteriormente clivadas, expondo o sítio ativo e as tornando ativas (VERHAMME; LEONARD; PERKINS, 2019). Em relação ao pró-peptídeo, a retirada dessa região nas proteínas que o possuem causa um impedimento na construção da sua correta estrutura 3D ativa, se estiverem desnoveadas. Isso aplica-se tanto para proteínas maduras desnaturadas *in vitro* quanto para proteínas sintetizadas em sistemas de expressão artificial sem a região pró, por exemplo. Tudo isso indica o alto nível de controle desse tipo de regulação (DEMIDYUK et al., 2010).

Essa família tem uma grande abrangência e ultimamente foram publicados inúmeros trabalhos demonstrando sua importância tanto na fisiologia normal quanto em patologias humanas (DANA; PATHAK, 2020). Catepsinas são descritas em humanos atualmente com 11 representantes (B, C, F, H, L, K, O, S, V, W, X, Z) dos quais a maioria é amplamente expressa nos tecidos humanos. Dentre estas, as catepsinas B e L apresentam maior destaque, visto que apresentam proteínas homólogas em diferentes organismos (VIZOVIŠEK; FONOVIC; TURK, 2019). Ambas catepsinas já foram descritas em diversas células e, portanto, as suas atuações intra e extracelulares foram largamente investigadas, notando-se os papéis destas proteases em múltiplos processos fisiológicos e patológicos (COULOMBE et al., 1996). Essas peptidases foram igualmente correlacionadas à progressão de câncer de mama, como também à taxa de reincidência do mesmo após tratamentos. Ao que concerne outros tipos de câncer, foi possível relacionar suas concentrações à probabilidade de sobrevivência dos pacientes no estudo de tumores pulmonares, cerebrais e de cólon, por exemplo (YADATI et al., 2020).

A catepsina L é produzida na forma de pré-pró-peptídeo, onde ela é endereçada ao lisossomo para que, então, ela possa modificar sua conformação para a forma madura da proteína com a inserção no pH baixo encontrado na organela. Seu envolvimento em doenças foi observado por meio da sua relação com diabetes, fibrose hepática, disfunção renal e diversos cânceres invasivos, onde ela adquire funções extremamente desreguladas (DANA; PATHAK, 2020). Partindo para outros organismos com o foco como alvo terapêutico, a catepsina L apresenta ainda notável atuação em tripanosomatídeos, nos quais ela é amplamente expressa e possui grandes semelhanças bioquímicas com a catepsina B de *T. cruzi*, dispondo inclusive de inibidores iguais (PEREIRA et al., 2019). Conhecida como brucipaina ou rodesaina em *Trypanosoma brucei*, ela é responsável, junto com outras cisteíno proteases, pela conclusão do ciclo biológico do parasito no hospedeiro e, em vista disso, foi observado que sua inibição é capaz do prolongamento da vida de camundongos infectados (ABDULLA et al., 2008). Estudos mais recentes mostraram ainda que essa protease está relacionada com a facilitação da entrada do vírus SARS-COV-2 em células humanas, assim como sua concentração circulante está correlacionada com a progressão e severidade da COVID-19 (ZHAO et al., 2020).

### 1.6.1 Catepsina B

A catepsina B é outro exemplo de uma participante muito relevante das cisteíno proteases e está inserida dentro da família das papaínas. Ela é sintetizada como pré-pró-enzima em forma de zimogênio e, em humanos, é constituída por 339 aminoácidos com o peso molecular de aproximadamente 30 kDa em sua forma madura e ativa (CAVALLO-MEDVED; MOIN; SLOANE, 2011). Caracteriza-se como uma cisteíno protease lisossomal, assim como a catepsina L, e apesar de apresentar homologia funcional e estrutural com outras enzimas da mesma família, exibe características que a diferem das demais catepsinas: a posição que a ligação peptídica é cortada, uma vez que ela possui atividade dupla de endo e exopeptidase, como também a presença do *loop* de oclusão em sua conformação que oculta o sítio ativo impedindo a ligação do substrato anteriormente à maturação da enzima (NÄGLER et al., 1997). Suas atuações são bastante estudadas em humanos, visto que sua cristalografia em células hepáticas foi realizada ainda em 1991 exibindo sua estrutura tridimensional, onde foi possível constatar a localização do sítio ativo, do *loop* de oclusão e suas cadeias, como observado na **Figura 5** (MUSIL et al., 1991). A protease possui um sítio ativo com uma tríade catalítica em forma de “V” composto pelos resíduos de cisteína (Cys), histidina (His) e asparagina (Asn). Em humanos, estes resíduos se encontram nas posições Cys108, His278 e Asn298, quando ainda há a presença do pró-peptídeo, e Cys29, His199 e Asn219, em sua forma madura. A proteína forma 6 pontes de dissulfeto e possui ainda 2 resíduos de cisteína que não são pareados, a presente no sítio ativo é a Cys240 (SCHMITZ et al., 2019).



**Figura 5. Modelo da estrutura 3D da catepsina B humana madura (PDB 1HUC) (MUSIL et al., 1991).** Figura gerada pelo programa UCSF CHIMERA, destaque aos resíduos do sítio catalítico em vermelho e o *loop* de oclusão em azul.

O processo de maturação da catepsina B humana ocorre devido à remoção da porção N-terminal pró-peptídica, a extensão C-terminal e a ligação dipeptídica entre os resíduos que interagem com o *loop* de oclusão, resultando assim na forma madura e ativa (TURK et al., 1996). A catepsina B humana primeiramente é produzida em sua forma precursora e inativa de aproximadamente 45kDa ainda contendo a presença do peptídeo sinal nos primeiros 17 resíduos de aminoácidos, o qual a endereça para o lúmen do retículo endoplasmático rugoso (RER), onde ela é sintetizada, para formar a pró-catepsina resultando em uma conformação da enzima de aproximadamente 33 kDa. Então, esta forma é carregada para o aparato de Golgi através do RER, onde sofre glicosilação em 2 resíduos de asparagina ao entrar em contato com oligossacarídeos contendo manoses fosforiladas. Por intermédio da glicosilação, a protease tem a capacidade de se ligar a receptores manose-6-fosfato na via do trans-Golgi, possibilitando o transporte via vesículas para os lisossomos, a qual é finalmente convertida para a conformação ativa e madura composta de 2 subunidades, uma cadeia pesada com 25 kDa e uma cadeia leve com 5 kDa, mediante à dissociação da região pró. A pró-catepsina tem a importante função dupla de ser auto estabilizadora e inibidora para que não ocorra a maturação precoce da enzima anteriormente à chegada aos lisossomos. Este processo de maturação é pH dependente e a atividade catalítica da protease é dependente da conformação do *loop* de oclusão (MIJANOVIĆ et al., 2019; SLOANE, 2014).

O processamento do pró-peptídeo, para causar a maturação da enzima, ocorre pela presença de pH baixo, aproximadamente 4,5, em que as interações entre a região pró e a porção madura se enfraquecem ocasionando uma conformação mais aberta e móvel. O pró-peptídeo apresenta predominantemente um domínio  $\alpha$ -hélice que fica posicionado como um “gancho” no topo do sítio catalítico, onde interage com a parte madura e fortalece esta interação. A cadeia gerada pela região pró serve como um elo entre a região N-terminal do segmento maduro e o “gancho” formado, pois se estende através da fenda do sítio ativo em direção a porção N-terminal da enzima madura, oposta ao sítio de ligação do substrato. A conformação obtida após a inserção em pH neutro desencadeia a clivagem intermolecular do pró-peptídeo, devido à atividade catalítica presente na pró catepsina B que, apesar de ser baixa, é suficiente para promover ativação auto catalítica do zimogênio, preferencialmente entre os resíduos Met56-Phe57 (PUNGERČAR et al., 2009).

A função de bloqueio exercida pelo *loop* de oclusão é explicada pela interação que os resíduos da estrutura possuem com as cadeias laterais da proteína. Em humanos, a histidina 110 (His110) e arginina 116 (Arg116) presentes no *loop* realizam interações de pontes de sal com os resíduos de ácido aspártico (Asp22) e ácido aspártico 224 (Asp224), respectivamente. Dessa forma, sua estrutura se torna mais compacta criando um impedimento físico no acesso ao sítio catalítico (NÄGLER et al., 1997). Portanto, esta região que dispõe de aproximadamente 20 aminoácidos, por meio desse evento, limita o espaço de interação para 2 resíduos de aminoácidos da região C-terminal da ligação peptídica do substrato. Assim, evitando a ligação de substratos grandes e, conseqüentemente, favorecendo a atividade de exopeptidase à proteína. A atividade de exopeptidase é determinada pela interação polar dos resíduos protonados de histidina (His110 e His111) presentes no *loop* permitindo a existência de um espaço acceptor apropriado para ligação com o grupo carboxil C-terminal do substrato. Logo, a protease possui uma atividade de carboxidipeptidase, principalmente observada em pH ácido. Quando a enzima se localiza nos endossomos, ocorre a mudança para pH neutro, resultando em estados de protonação e desprotonação dos resíduos formadores das pontes de sal. O *loop*, por conseqüência, assume outra conformação, não mais bloqueando o sítio catalítico, proporcionando a entrada de substratos maiores favorecendo assim a atividade de endopeptidase. Estudos referentes à mutação e/ou deleção dos resíduos pertencentes ao *loop* indicaram a formação também de uma estrutura mais aberta

viabilizando a atividade endopeptídica, visto que a protease perde sua atividade exopeptídica (SLOANE, 2014; SCHMITZ et al., 2016; SCHMITZ et al., 2019).

Desde a resolução de sua estrutura, os aspectos voltados a suas funções e atuações foram explorados culminando no foco do seu papel em múltiplas patologias, principalmente câncer, que foram observados por meio de seus níveis de expressão (SLOANE, 2014). Durante a investigação sobre a peptidase, notou-se que a sua regulação durante a transformação maligna das células, ou seja, a formação de novas células tumorais, se torna alterada fazendo com que ocorra a sua superexpressão, diferentemente da fisiologia normal na qual ela se encontra extremamente regulada. Esta alta expressão já foi descrita em câncer pulmonar, colorretal, cerebral, mamário e prostático (RAO, 2013). Outro dado interessante é que células tumorais expressam transcritos variantes de catepsina B e L usando promotores e *splicings* alternativos. Estes transcritos variantes de catepsina B foram identificados altamente expressos em estudos referentes a gliomas e foram submetidos a infrarregulação, a qual resultou em redução de crescimento, invasão e da angiogênese do tumor (MOHAMED; SLOANE, 2006; RAO, 2003).

Além disso, há indícios de que a atividade carboxidipeptídica da protease está envolvida na taxa de proliferação celular, o que pode se tornar um problema em pessoas que tem a neurogênese comprometida, como na doença de Alzheimer (DA) e Huntington, por exemplo (MIJANOVIĆ et al., 2019). Os níveis de CatB são aproximadamente 50% maiores no soro de pacientes com DA comparados com pacientes controles, o que posteriormente foi correlacionado também com disfunção cognitiva e demência. Estes níveis plasmáticos estão presentes tanto em indivíduos com casos leves quanto graves de DA, mas estes graus crescem a medida que a doença progride (SUN et al., 2015). Ainda, a catepsina B apresenta localização extracelular anormal no cérebro de pessoas com DA, contrastando com a localização intracelular nos lisossomos comumente exibida, o que indica uma redistribuição da proteína nos cérebros dos afetados pela doença (HOOK et al., 2020; ZHANG et al., 2005).

### 1.6.2 Catepsina B de protozoários

As cisteíno proteases, como a catepsina B, são igualmente descritas em protozoários causadores de doenças e desempenham papéis importantes no ciclo de

vida desses organismos (SAJID; MCKERROW, 2002). Em relação a suas atuações em *T. brucei*, o mRNA da protease é suprarregulado na forma sanguínea do parasito, indicando que ela está envolvida no sistema de degradação de proteínas do hospedeiro. O nocaute simples da catepsina B apresentou diminuição na taxa de crescimento e morfologia anormal. A utilização de RNAi com o objetivo de nocaute duplo do gene *in vitro* indicou que o compartimento do lisossomo/endossomo alargueceu, contribuindo para uma morfologia disfórmica do parasito com falha no processo de citocinese. Estudos *in vivo* de RNAi mostraram que camundongos obtiveram alta taxa de sobrevivência após dois meses de infecção com ausência de parasitos na corrente sanguínea e de esplenomegalia. As pesquisas apontaram baixa viabilidade e grande letalidade para o parasito (MACKEY et al., 2004; ABDULLA et al., 2008; O'BRIEN et al., 2008).

Testes *in vivo* infectando camundongos C57BL/6 deficientes de algumas cisteíno proteases com *Leishmania major* revelaram que apenas os animais deficientes de catepsina B exibiram maior resistência à infecção com melhora nas lesões e diminuição da carga parasitária. Ainda, a inibição da protease modulou uma resposta celular Th1 protetiva precoce em camundongos BALB/c, que está provavelmente associada a intrínseca capacidade das células com deficiência da proteína de secretarem mais interferon gama (IFN $\gamma$ ) e reduzirem a proliferação por parte das células TCD3 com uma resolução mais rápida da inflamação (GONZALEZ-LEAL et al., 2014; RASID et al., 2016).

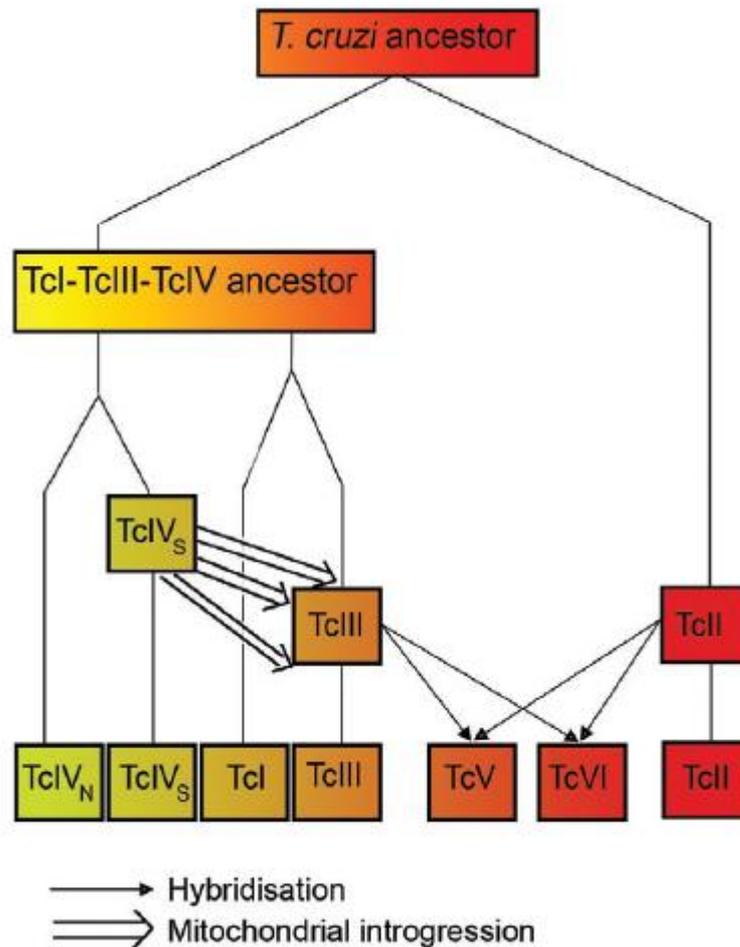
Em se tratando de *T. cruzi*, a proteína foi pouco elucidada, porém a protease possui alta similaridade com catepsinas B de outros organismos, como por exemplo de *L. mexicana*, *Caenorhabditis elegans* e a humana (CTSB), apresentando muitas regiões conservadas, principalmente na porção N-terminal das proteínas. Porém, a protease mostrou pH ótimo de 4,5, atividade enzimática em substratos proteicos e sensibilidade para inibidores parecidos com a cruzipaína, ou catepsina L de *T. cruzi*. (GARCIA et al., 1998). Então, a porção madura da catepsina B de *T. cruzi* foi caracterizada com 30 kDa que é produzida em todas as formas do parasito. A superexpressão da proteína atribuiu a epimastigotas uma maior capacidade replicativa e de transformação em tripomastigostas metacíclicos comparada com o genótipo selvagem (NÓBREGA, 2001). Nota-se, porém, que a expressão da protease em epimastigotas é superior a outras formas de desenvolvimento do parasito, assim como apresenta a atividade hidrolítica de modo principal sobre gelatina em condições

acídicas, indicando capacidade de clivar proteínas do hospedeiro quando se encontra no meio extracelular. Foi possível também encontrar sua localização em grânulos lisossômicos na porção posterior de epimastigotas e por todo o citoplasma em amastigotas (RESTREPO, 2001).

### 1.7 Genética e biologia molecular do *Trypanosoma cruzi*

No que concerne a genética de *T. cruzi*, o parasito possui o seu DNA compartimentalizado no núcleo, ao mesmo tempo que apresenta dentro de sua mitocôndria, o DNA mitocondrial, ou também chamado de kDNA, localizado dentro da estrutura chamada cinetoplasto. O kDNA é organizado em mini e maxi círculos concatenados, que estão igualmente presentes em outros tripanossomatídeos, compreendendo aproximadamente 15% do DNA total da célula. Os maxi círculos são moléculas grandes análogas ao DNA mitocondrial de outros eucariotos com baixo número de cópias e codificam os genes estruturais para a correta função mitocondrial. Já os mini círculos são moléculas pequenas com alto grau de conservação e múltiplas cópias responsáveis pela replicação e edição dos transcritos mitocondriais exclusivas de tripanossomatídeos. As diferenças fenotípicas entre cepas de *T. cruzi* ocorrem pela diferença na organização e conteúdo do DNA e kDNA (JUNQUEIRA; DEGRAVE; BRANDÃO, 2005; TEIXEIRA; NASCIMENTO; STURM, 2006).

Por possuir alta variabilidade genética, as linhagens do parasito são divididas em 6 DTUs, ou unidades discretas de tipagem, que abrangem todas as cepas existentes. As DTUs são categorizadas com números romanos de I-VI abrigando as cepas que possuem maior relação genética entre si, visto na **Figura 6** (HIGUERA; GUHL; RAMÍREZ, 2013). A cepa de referência para o *T. cruzi* é a CL Brener, por se tratar de uma cepa experimentalmente bem caracterizada, que teve seu genoma sequenciado em 2005. Ela se caracteriza como híbrida, em virtude da presença dos alelos Esmeraldo-Like e Non-Esmeraldo-Like originários de diferentes ancestrais com DTUs distintas (TcII e TcIII, respectivamente), portanto se enquadrando no DTU VI, a qual engloba as linhagens resultantes de trocas genéticas entre as DTUs II e III (EL-SAYED, 2005).

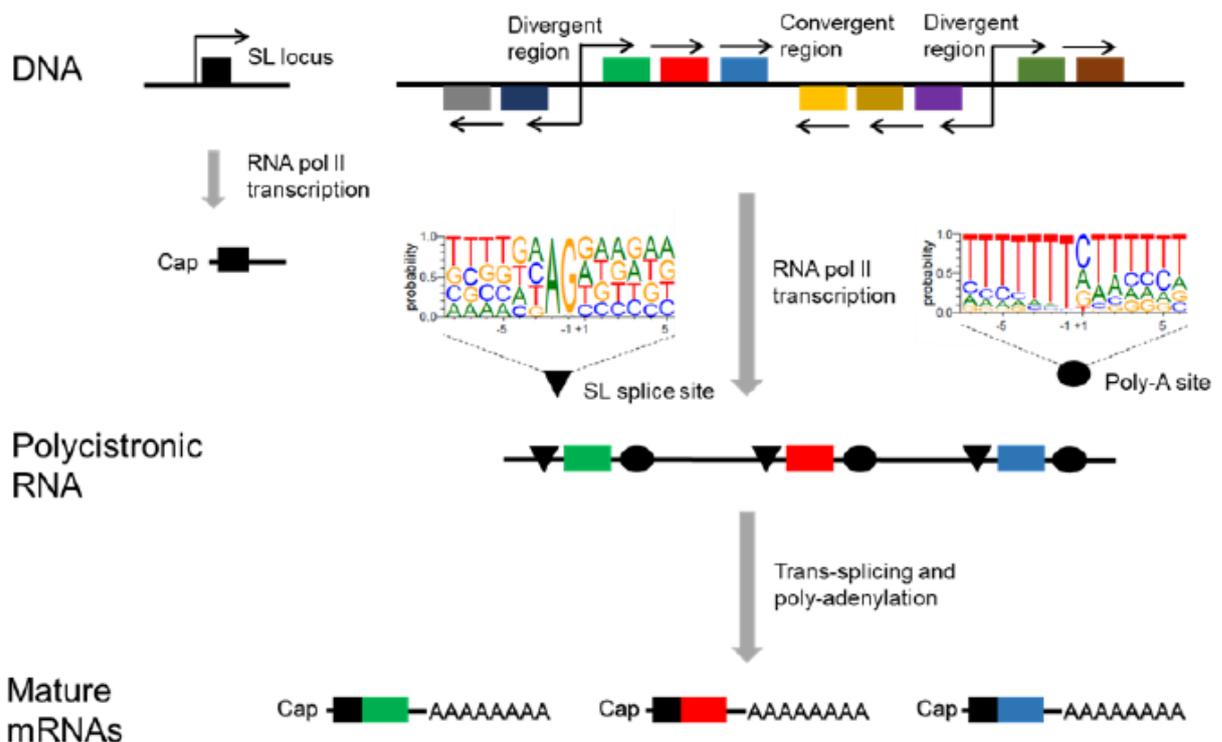


**Figura 6. Representação da classificação filogenética das DTUs I-VI de *T. cruzi* indicando os ancestrais em comum e as formações de cada DTU. Adaptado de (TOMASINI; DIOSQUE, 2015).**

Em comparação a outros eucariotos, o evento de transcrição do *T. cruzi* ocorre em um processo diferente da maioria (**Figura 6**). Da mesma maneira que outros tripanossomatídeos, os genes codificadores de proteína presentes no *T. cruzi* são organizados em *clusters* não sobrepostos na mesma fita de DNA contendo longas unidades policistrônicas de transcrição de genes. Os *clusters* não possuem tamanhos determinados podendo até ocupar o cromossomo inteiro, assim como os genes inseridos neles não são necessariamente relacionados ou possuem funções parecidas, sendo apenas separados por pequenas regiões intergênicas (VAZQUEZ, 2007). Estas regiões correspondem a espaços não codificantes de proteína, mas que são igualmente transcritos. Dentro dessas regiões são encontradas sequências fundamentais que guiam as etapas envolvidas na regulação da expressão gênica

como os trechos de polipirimidinas e a sequência de nucleotídeos-sinal AG (NOZAKI; CROSS, 1995).

A via do promotor RNA polimerase II para a iniciação da transcrição não é identificada em *T. cruzi* e, conseqüentemente, o processo não é controlado no nível de genes individuais e sim do *cluster* completo. A transcrição ocorre bidirecionalmente e o pré-mRNA formado contém todos os genes presentes naquele *cluster*. Assim, ele está apto para passar pela etapa de processamento que consiste em 2 passos: o capeamento da região 5' e a poliadenilação. Primeiramente, a sequência líder (SL) é transcrita e produz uma região CAP, comportando 39 nucleotídeos, que se liga a região 5' de cada mRNA guiado pela sequência de polipirimidinas presentes na região intergênica e detém a função de estabilizar este mRNA. Como resultado, o mRNA é excisado propiciando assim, a poliadenilação. O segundo passo envolve a inserção de uma sequência na região 3' do mRNA contendo múltiplas adeninas que são similarmente guiadas pelas sequências presentes nas regiões intergênicas, resultando assim nos mRNAs maduros individuais (CLAYTON, 2019; HERREROS-CABELLO et al., 2020).



**Figura 7. Representação do mecanismo de processamento pós- transcricional do mRNA em *T. cruzi*.** Adaptado de (HERREROS-CABELLO et al., 2020).

## 1.8 Utilização do sistema CRISPR/Cas9

Na procura de métodos eficientes para edição genética de tripanossomatídeos e, principalmente, o *T. cruzi*, o CRISPR/Cas9 (*Clustered Regularly Interspaced Short Palindromic Repeats-associated with gene Cas9*) surgiu como uma ferramenta promissora. O método foi desenvolvido com base no sistema imunológico procariótico, específico de algumas bactérias, voltado para a defesa contra DNA exógeno eficientemente alvejando sequências virais e plasmidiais (LANDER; CHIURILLO; DOCAMPO, 2016).

Antes da descoberta do seu funcionamento, em 1987, a presença de uma sequência contendo 29 nucleotídeos que se repetia no decorrer do genoma intercalada por sequências curtas não repetitivas e não relacionadas (posteriormente chamadas de espaçadores) foi observada em *Escherichia coli*. Em seguida, outros organismos foram descritos como detentores desta sequência. Até que então, esses nucleotídeos constantes foram identificados como CRISPR (GUPTA et al., 2019). Uma das primeiras constatações da função biológica do CRISPR aconteceu apenas em 2005 quando foi observado que as aparentes sequências repetidas que intercalavam os seguimentos idênticos CRISPR obtiveram homologia a DNA invasivo viral e plasmidial. Rapidamente o processo foi associado ao conhecido mecanismo de RNA interferente (RNAi) exibido em eucariotos, porém os componentes da maquinaria não possuíam homólogos com as proteínas Cas. O RNAi, em eucariotos, consiste em uma metodologia adaptada da resposta de defesa natural à invasão de fitas duplas de RNA ocasionadas geralmente por vírus. A técnica tem a capacidade de originar fenótipo de mutantes nulos aos organismos com o gene silenciado, porém, em vez de causar a eliminação do gene, causa apenas a sua infrarregulação. Posteriormente, com o advento do sequenciamento de diversos organismos, o sistema CRISPR/Cas foi altamente explorado até que os modelos adaptados foram caracterizados em condições nativas principalmente em *Streptococcus thermophilus*, *Streptococcus pyogenes* e *Pseudomonas aeruginosa* (BARRANGOU; HORVATH, 2017; BALAÑA-FOUCE; REGUERA, 2007).

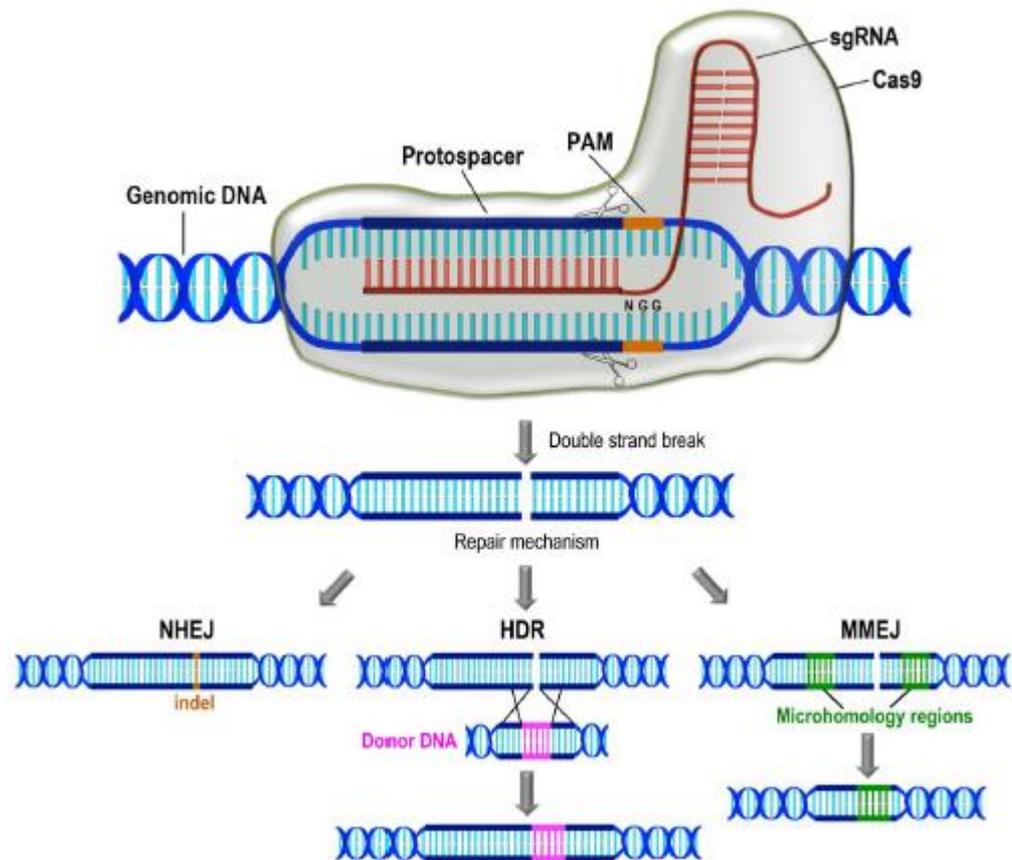
Atualmente, o sistema CRISPR/Cas é dividido em 2 classes que são subdivididas em 6 tipos e 19 subtipos. O CRISPR/Cas Classe 1 utiliza complexos efetores multi-proteínas, por outro lado o CRISPR/Cas Classe 2 opera com efetores

advindos de uma única proteína. Inserida na Classe 2, a endonuclease *SpCas9* de *S. pyogenes* é umas das mais utilizadas associadas com o sistema CRISPR para edição gênica. As aplicações do sistema associado com a *SpCas9* são imensuráveis. Entre os empregos mais utilizados da técnica estão: edição de modulação, transcrição e epigenética; *screening* em escala genômica; terapia em células e embriões; edição de bases (ZHANG; QUAN; WANG, 2018).

A edição de bases é muito utilizada em tripanossomatídeos, especialmente em *T. cruzi*, pois os organismos possuem alto grau de plasticidade genômica, ou seja, a tentativa de produção de mutantes nulos por outras abordagens geralmente provoca a criação de novas cópias dos alelos referentes ao gene de interesse, resultando assim, em mais rodadas de transfecção. A introdução do CRISPR/Cas9 para esse tipo de edição gerou grandes avanços na área. Em vista disso, a metodologia possui a *SpCas9*, uma endonuclease guiada por RNA, que na adaptação da tecnologia do CRISPR, pode ser guiada por uma sequência curta e específica de aproximadamente 20 pb chamada de RNA guia (sgRNA). O sgRNA contém uma sequência complementar ao gene alvo que necessariamente deve preceder um segmento PAM (motivo protoespaçador adjacente) o qual consiste em uma região repetida ao longo do genoma que marca o local de ligação à fita de DNA. Em CRISPR/Cas9, o PAM mais utilizado, para a localização, são os nucleotídeos: 5'-NGG. Desse modo, o sgRNA e a *SpCas9* formam um complexo ribonucleoproteico, o qual localiza o motivo PAM, ocasionando assim, a quebra da fita dupla do DNA. (RICO et al., 2018; YAGOUBAT et al., 2020).

Prontamente, o sistema de reparo de DNA é ativado e, em tripanossomatídeos, preferencialmente, o método utilizado é o de reparo por micro homologia (MMEJ). O mecanismo MMEJ consiste na recombinação de sequências homólogas curtas adjacentes ao corte, presentes em ambas as fitas de DNA, resultando em uma deleção de alguns nucleotídeos. Outro recurso de reparo é por junção de extremidades não homólogas (NHEJ) que apenas insere e exclui randomicamente nucleotídeos para a restauração do DNA, porém este dispositivo não é descrito em *T. cruzi* (**Figura 8**). E por último, recombinação homóloga direta (HDR), a qual usa um DNA doador para o reparo homólogo. No intuito de se evadir o sistema MMEJ, o uso de um DNA doador exógeno é estabelecido para que aconteça a correta disrupção do gene (GLOVER, 2011). Ainda, o emprego de cassetes que conduzem a homologia

direta com o objetivo de assegurar a eficiência do nocaute institui uma estratégia usual em *T. cruzi* (LANDER; CHIURILLO; DOCAMPO, 2016).



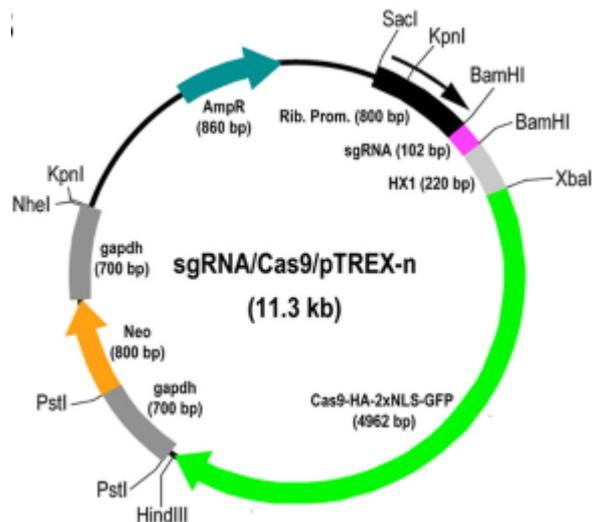
**Figura 8.** Representação do sistema CRISPR/Cas9 atuando na quebra da dupla fita de DNA dispondo dos componentes moleculares e dos mecanismos de reparo de DNA. Adaptado de (LANDER; CHIURILLO; DOCAMPO, 2016).

### 1.8.1 Edição gênica em *Trypanosoma cruzi*

No que diz respeito à edição gênica em *T. cruzi* utilizando o sistema CRISPR/Cas9, é necessária atenção a alguns fatores que podem influenciar no resultado. A endonuclease *SpCas9* já foi descrita como tóxica para *T. cruzi* e *Leishmania spp.*, o que futuramente pode atrapalhar sua correta expressão e atuação, portanto a forma como ela vai ser expressa deve ser bem avaliada (LANDER; CHIURILLO; DOCAMPO, 2016). Assim como a presença de *off-targets*, ou seja, regiões diferentes do gene de interesse que o *sgRNA* pode identificar e clivar, mesmo se tratando de uma sequência específica, o que impede a precisa supressão do alvo (LANDER et al., 2016). Entretanto, mesmo com estes empecilhos, o CRISPR/Cas9 em *T. cruzi* é caracterizado como uma ferramenta extremamente eficiente no nocaute,

exibindo 70% da população com fenótipo mutante, e rápida pois, após 2 dias da transfecção, é possível constatar evidentes reduções no níveis de expressão proteica (PENG et al., 2015).

Uma das ferramentas genéticas mais eficazes presentes em *T. cruzi* é a superexpressão de genes com o vetor pTREX (VAZQUEZ; LEVIN, 1999), um sistema que possibilita expressão constitutiva de genes guiada por um promotor de genes ribossomais, como também a integração da sequência no locus do DNA ribossomal. Uma das estratégias para aplicação do CRISPR utiliza o plasmídeo sgRNA/Cas9/pTREX-n (Figura 9) que possui o sgRNA e a *SpCas9* clonados em sua sequência a montante e a jusante, respectivamente, do sinal de *trans-splicing hx1*. O gene *hx1* é proveniente de regiões intergênicas de *T. cruzi* e caracteriza um sinal altamente eficiente para processamento de mRNA e direcionamento de expressão gênica. A *SpCas9*, presente no plasmídeo, está fusionada a um sinal exógeno de localização nuclear (*NLS*) e à *GFP* (proteína fluorescente verde). Dessa forma, os mutantes podem ser visualizados por citometria de fluxo com o núcleo marcado de verde, por exemplo. Ainda dispõe da marca de seleção com o antibiótico neomicina para auxílio na cultura celular. A maior vantagem do uso do sgRNA/Cas9/pTREX-n é a possibilidade de uma única rodada de transfecção para o nocaute de genes em *T. cruzi* (DOCAMPO, 2015; VAZQUEZ; LEVIN, 1999).



**Figura 9. Representação do plasmídeo sgRNA/Cas9/pTREX-n indicando a presença de todos os seus componentes.** O sgRNA, na cor rosa, flanqueado pelo promotor ribossomal e o sinal *hx1* com os sítios de restrição para *Bam*HI. Além da *SpCas9* fusionada à *GFP* e *NLS* como marca de seleção para neomicina. Adaptado de (DOCAMPO, 2015).

### 1.9 Modelagem da estrutura tridimensional da catepsina B de *Trypanosoma cruzi*

Tendo em vista uma melhor elucidação da função e das interações da catepsina B de *T. cruzi* com inibidores, a obtenção de sua estrutura tridimensional é imprescindível. Entretanto, é evidente a discrepância entre o número de sequências primárias depositadas em plataformas como o UniProt e o número de estruturas tridimensionais resolvidas no PDB. Por exemplo, até 2017, no UniProt havia mais 85 milhões de sequências primárias depositadas contra aproximadamente 120 mil estruturas 3D no PDB, indicando que apenas cerca de 0,2% das estruturas das proteínas chegam a ser resolvidas (ZHANG, 2018). Essa divergência pode ser explicada primeiramente pelos equipamentos usados, geralmente com o uso de técnicas como cristalografia por raio-X e espectroscopia por ressonância magnética nuclear, que elevam bastante o tempo e os custos relacionados à resolução experimental e cristalização da estrutura de proteínas, conseqüentemente, dificultando bastante o processo (ZHENG et al., 2019). Além disso, a obtenção do cristal da proteína é um processo laborioso. Os resultados obtidos devem ser validados e refinados e, mesmo com pessoas e maquinaria qualificadas, configurações discordantes da proteína ainda podem ser alcançadas, posto que em solução a proteína pode dispor uma conformação diferente da achada em cristal (ACHARYA; LLOYD, 2005). Para auxiliar nessa questão, a aplicação de estratégias computacionais tem sido essencial para preencher esta lacuna, gerando progresso e sofisticação no resultados obtidos (YANG et al., 2014). A modelagem da estrutura tridimensional da proteína possibilita a sua exploração em nível molecular, as interações com ligantes, assim como o conhecimento de sua estrutura quaternária, que proporciona uma maior compreensão dos sistemas biológicos em que a proteína está inserida e como ela o influencia (WATERHOUSE et al., 2018).

Diferentes abordagens computacionais podem ser utilizadas quando se trata do desenvolvimento do modelo tridimensional de uma proteína. Uma delas é a modelagem por homologia, que se baseia no princípio de que sequências primárias idênticas apresentam homólogos estruturais conservados, viabilizando assim, a modelagem baseada em um *template* ou molde. O método consiste em modelar interações não caracterizadas usando estruturas experimentais de protômeros homólogos interativos como o molde (BERTONI et al., 2017). Ou seja, a predição do modelo acontece a partir da estrutura 3D, já resolvida e depositada em banco de

dados, de uma outra proteína, que possivelmente é relacionada evolutivamente com a proteína de interesse. Para tal, essas sequências depositadas precisam ter no mínimo 25% de identidade com a sequência de aminoácidos do alvo. Além disso, o limite de predição do modelo está intimamente correlacionado com a similaridade entre o alvo e o molde (VERLI, 2014). Um exemplo de servidor que faz modelagem por homologia é o servidor Swiss Model (WATERHOUSE et al., 2018). A partir de uma dada sequência primária, o programa busca por proteínas resolvidas experimentalmente no banco de dados de proteínas (PDB) e gera um modelo baseado em estruturas resolvidas de sequências homólogas. Como resultado, ele classifica os PDB molde encontrados, por meio da porcentagem de identidade sequencial, porcentagem de cobertura dos resíduos da proteína no molde e dos valores estimados globais de qualidade, entre eles, o QMEAN (*Qualitative Model Energy Analysis*) que se trata de uma função composta de pontuação com base nos aspectos geométricos da estrutura de proteínas (BENKERT; TOSATTO; SCHOMBURG, 2008).

Outro método utilizado é o baseado no tipo de enovelamento ou *threading* da proteína. O conceito de *threading* está ligado ao processo de identificação de *templates* de proteínas com estruturas resolvidas experimentalmente em banco de dados de enovelamento (*Protein Folding Database*) que têm uma estrutura potencialmente similar a proteína alvo (ZHENG et al., 2019). A metodologia se baseia em estruturas 3D serem mais conservadas que sequências primárias, indicando que sequências com baixa identidade ainda são capazes de produzir estruturas altamente similares (VERLI, 2014). Um exemplo de servidor que utiliza essa metodologia é o I-Tasser. Este servidor utiliza o *threading* por meio de alinhamento múltiplo em PDBs, localizando assim, *templates* estruturais. O modelo atômico é construído com base em simulações de fragmentos contínuos estruturais excisados dos *templates* obtidos dos PDB moldes. Após, predições da função são derivadas da comparação do modelo gerado com proteínas de funções conhecidas (YANG; ZHANG, 2015). Como resultado, uma lista com 10 PDB moldes é elaborada e designa até 5 modelos da proteína alvo, aos quais atribui um valor determinante da confiabilidade do modelo chamado *C-score*, conferindo assim, uma pontuação de qualidade global estimada que varia de -5 a 2 (ZHENG et al., 2019). O programa permite também a indicação de algum PDB de sua escolha para que seja obrigatoriamente incluído na busca.

Por fim, o método baseado em aprendizado de máquina (AM), que incorpora conhecimento físico e biológico da estrutura da proteína, realizando alinhamentos de

múltiplas sequências, gerando modelos com alta acurácia. AM pode ser utilizado para predição de estruturas 3D de proteínas que não possuem homólogos em bancos de dados, por exemplo, com precisão. Os algoritmos viabilizam a predição das orientações e distâncias entre os resíduos de aminoácidos definindo uma estrutura tridimensional otimizada. Um exemplo é o servidor trRosetta (YANG et al., 2020) que aplica redes neurais profundas e modelagem por homologia para construir o modelo. Isso induz o modelo a ter características gerais de estruturas de proteínas conhecidas, gerando assim, uma alta probabilidade de o resultado refletir proteínas “ideais” que ocorrem naturalmente. Dos parâmetros que avaliam a qualidade do modelo gerado, o TM-score predito representa a estimativa de confiança do modelo, variando de 0 a 1, sendo TM-score > 0,6 um modelo com topologia predita corretamente.

Seguindo este mesmo método, o programa AlphaFold (JUMPER et al., 2021) também é um exemplo de emprego da ferramenta computacional que possibilita a modelagem com acurácia em nível atômico ainda que não haja estrutura similar resolvida. Apesar de também operar com AM, o AlphaFold faz uso de outro tipo de rede neural para formar a estrutura tridimensional da proteína que é dividido em 2 etapas. Primeiramente, o dado de entrada é processado construindo diversas camadas de sequências alinhadas separando-as em blocos de interesse e não interesse, de acordo com a interpretação da rede. Posteriormente, é produzida a estrutura tridimensional conforme a translação e rotação de cada resíduo da proteína obedecendo as relações espaciais e evolucionárias. Como parâmetro, o programa avalia o resultado da estrutura e atribui scores de confiabilidade para cada resíduo elaborado com pontuação específicas variando de 0 a 100.

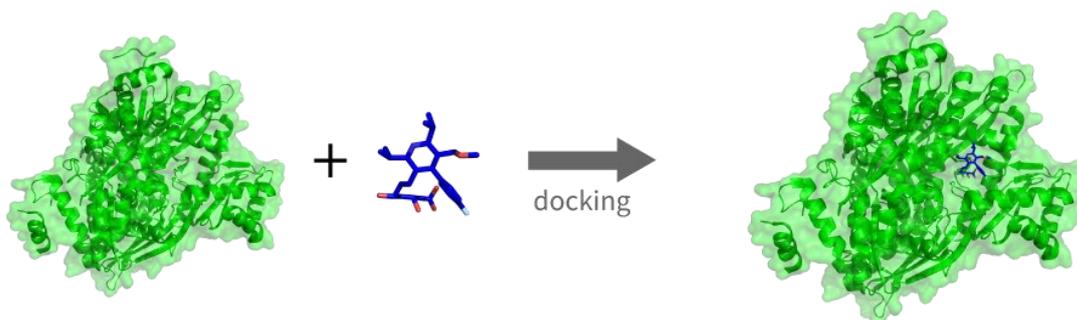
Todos os parâmetros de qualidade avaliados para os modelos foram detalhados na metodologia do trabalho.

### 1.10 Triagem virtual baseada em *docking* molecular

Com o advento de novas estruturas de proteínas nos últimos anos, ocorreu, do mesmo modo, o aumento de pesquisas voltadas tanto para o esclarecimento de suas funções quanto para o desenvolvimento de fármacos. E, assim, metodologias computacionais se tornaram fundamentais para o progresso das diferentes aplicações existentes, como por exemplo, na descoberta e otimização de protótipos a fármacos

e, também, no estudo da interação proteína-ligante (KITCHEN et al., 2004). E assim, por exemplo, a triagem virtual de compostos foi capaz de avançar cada vez mais, ajudando no aumento de pesquisas na área de desenvolvimento de fármacos que alcançaram últimos estágios da fase clínica e ao mercado como medicamento. Uma das ferramentas que possibilitaram esse avanço foi a utilização da metodologia de *docking* (MOITESSIER et al., 2008).

O *docking* molecular tem a função de prever a conformação de ligantes no sítio da proteína e a afinidade de ligação entre um receptor e um ligante. Esta conformação geralmente é calculada manipulando ambas as estruturas não ligadas, ou seja, receptor e ligante separados, para então realizar o encaixe dos dois por parte do programa com o objetivo de mostrar as conformações pós-ligação e suas afinidades (**Figura 10**) (TROTT; OLSON, 2010). Apesar dos diferentes programas de *docking* possuírem especificidades, eles se baseiam em dois componentes principais. O primeiro denominado algoritmo de busca explorará a conformação espacial tanto do ligante quanto do receptor. O segundo é a função de pontuação (*score*) que avalia e classifica as poses de *docking*. Em geral, os *scores* refletem a energia de ligação entre ligante e proteína, quanto mais negativos, maior é a interação. Ambos os componentes são essenciais para os resultados do *docking*, uma vez que definem a acurácia da predição dos modos de ligação e a futura aplicação na otimização e triagem virtual (MOITESSIER et al., 2008).



**Figura 10.** Representação dos componentes e o resultado do *docking* com o encaixe de um receptor e um ligante.

Existem diferentes categorias utilizadas pelos programas para avaliar a flexibilidade dos ligantes. O primeiro é chamado de método sistemático e é feito com o objetivo de explorar todos os graus de liberdade da molécula. Este método pode ocorrer de duas formas distintas: o *docking* de vários fragmentos da molécula e a ligação covalente entre eles ou então por dividi-los em partes rígidas e flexíveis que geralmente são o centro do ligante e suas cadeias laterais, respectivamente, para

então realizar o *docking*. Incluso a este método, alguns programas empregam a ligação das estruturas rígidas através de interações químicas acessíveis no sítio ativo, como por exemplo ligações de hidrogênio, sítios aromáticos, hidrofóbicos, entre outros. Uma vantagem do método sistemático é que propicia um conhecimento melhor do espaço deste sítio, possibilitando uma noção da extensão e formato do sítio ativo, através de características estruturais que favorecem a interação do ligante. Outra metodologia utilizada pelos programas é o método randômico ou método estocástico que realiza mudanças aleatórias em um único ligante e/ou múltiplos ligantes. Dentro desse método podem ser usadas duas abordagens: Monte Carlo e algoritmos genéticos. Monte Carlo atua na energia do sistema gerando um *score* que calcula a distância das ligações com o sítio ativo fazendo com que se descubra motivos escondidos melhorando a ligação. O algoritmo genético é uma busca na qual os cromossomos codificam tanto a orientação quanto a conformação dos graus de liberdade do ligante, que são otimizados por operadores dos programas por *crossover* e mutações. O método randômico possibilita, então, a exploração das diversas conformações assumidas, definindo quais são aceitas ou não. Por último, o método de simulação por dinâmica molecular, no qual simula várias partes do sistema proteína-ligante em diferentes temperaturas em um determinado tempo em busca da melhor pose. Porém, este método encontra dificuldade em atravessar as barreiras de alta energia durante as simulações, acomodando os ligantes em locais com energia mínima (KITCHEN et al., 2004; HART; READ, 1992; WASZKOWYCZ; CLARK; GANCIA, 2011).

Um dos maiores problemas na aplicação do *docking* é como o programa lida com os diferentes graus de liberdade que as estruturas podem apresentar, visto que são moléculas flexíveis e torcionáveis em suas ligações internas, e assim, existe uma dificuldade por parte de alguns programas em otimizar e computar corretamente essas múltiplas conformações adotadas pelas moléculas. Evidenciando assim, uma grande limitação com relação à flexibilidade das macromoléculas. Outra dificuldade se apresenta com a baixa correlação entre *scores* e afinidade experimental. Em se tratando de *docking* receptor-ligante, o ideal é que o programa seja capaz de apurar ligantes grandes e extremamente flexíveis considerando simultaneamente a flexibilidade admitida pelo receptor (DE MAGALHÃES; BARBOSA; DARDENNE, 2004a).

### 1.10.1 Compostos inibidores

Compostos inibidores são muito estudados nas interações patógeno-hospedeiro, pois apresentam alta capacidade de causar intervenções em proteases que são importantes para a sobrevivência do organismo parasitário, por exemplo (BRYAN T. MOTT, RAFAELA S. FERREIRA, 2010). Tripanossomatídeos, por apresentarem importantes proteases homólogas, revelam a necessidade da elucidação das funções das proteínas que envolvem processos biológicos e patológicos. Uma das maiores metas dos estudos de função dessas proteases é o desenvolvimento de moléculas inibitórias que atuem, de preferência, em nível nanomolar com ligação covalente reversiva e de forma seletiva para a proteína de um organismo específico (SARTORI et al., 2019). Porém, a similaridade estrutural entre as proteínas facilita a escolha de potenciais compostos que interajam com os seus sítios ativos e, conseqüentemente, auxiliem nos testes de inibição e seletividade. Isso possibilita a identificação de regiões que gerem essa seletividade (FERREIRA, 2019).

Considerando-se que não há inibidores descritos para catepsina B de *T. cruzi*, buscou-se identificar compostos inibidores conhecidos dessa protease de outros organismos. Foram selecionados 14 compostos ativos de cinco artigos (CIANNI et al., 2021; JÍLKOVÁ et al., 2021; MALLARI et al., 2009; NDAO et al., 2014; SCHMITZ et al., 2016). Esses compostos apresentaram alto índice de inibição da catepsina B em *T. brucei* (TbCatB), *Schistosoma mansoni* (SMCB1) e *Homo sapiens* (CTSB) (**Figura 11**).

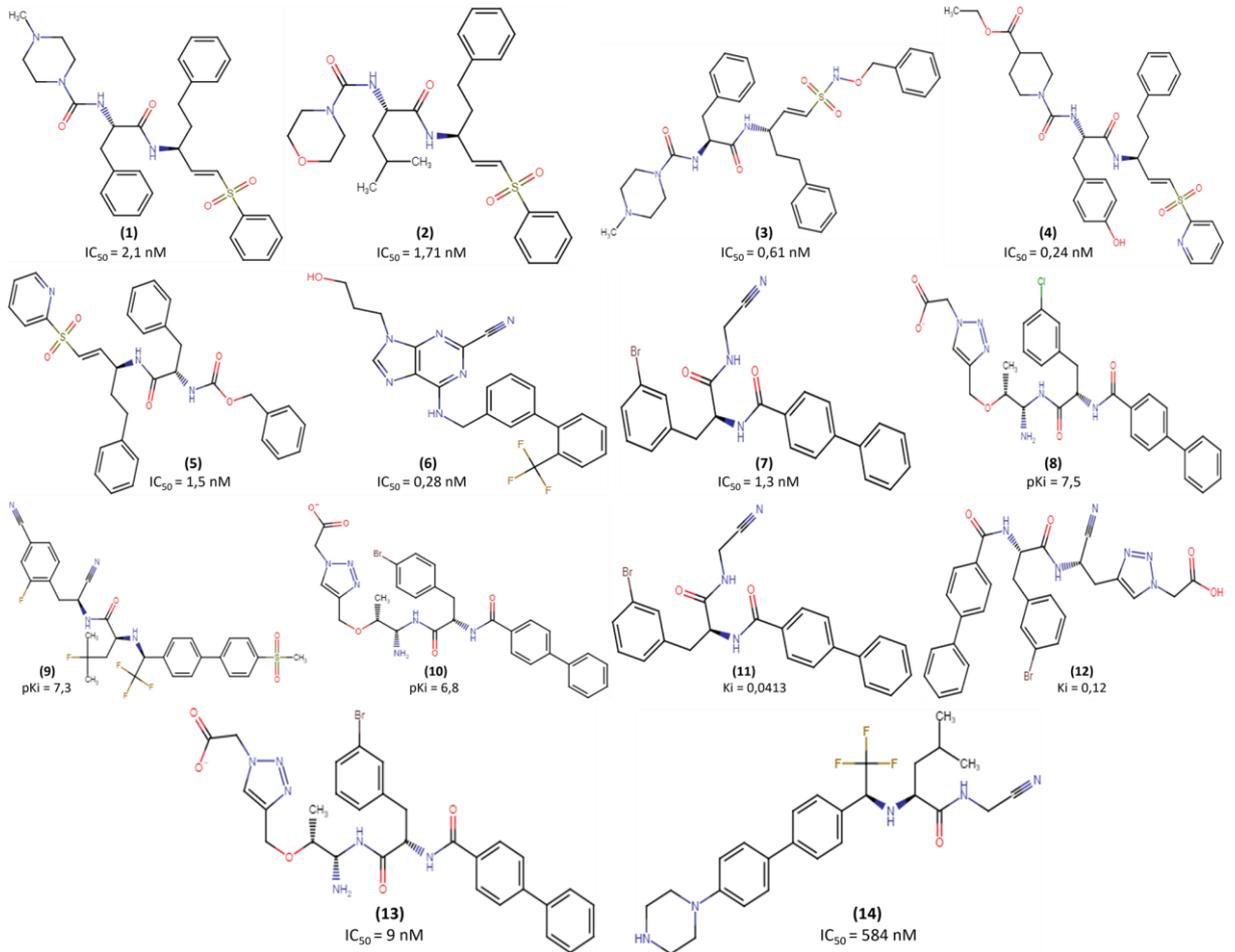


Figura 11. Estrutura 2D de compostos inibidores de catepsina B de *S. mansoni* (compostos 1-5) *T. brucei* (compostos 6 e 7), e *H. sapiens* (compostos 8-14).

# **JUSTIFICATIVA**

## 2 Justificativa

A doença de Chagas é uma enfermidade endêmica, que afeta milhões de pessoas em regiões tropicais, principalmente na América Latina. Trata-se de uma Doença Tropical Negligenciada (NTD) com escassas opções de tratamento farmacológico. Dos medicamentos aprovados entre os anos de 1975-2004, somente 1,3% foi destinado a doenças negligenciadas (CHIRAC; TORREELE, 2006), sendo apenas dois deles para a doença de Chagas. Além disso, a quimioterapia utilizada desde a década de 70 apresenta baixa eficácia, principalmente na fase crônica, e efeitos adversos severos. Devido a esses fatores, existe a extrema necessidade de buscar novos fármacos menos tóxicos, mais acessíveis e que sejam potenciais alvos quimioterápicos (GILBERT, 2017) é premente.

Em parasitos, cisteíno proteases ortólogas de mamíferos se mostraram amplificadoras do processo de invasão. As catepsinas B de parasitos são exemplos de cisteíno-proteases que desempenham papéis importantes no ciclo de vida desses organismos, podendo estar envolvidas na degradação de proteínas do hospedeiro. A catepsina B de *T. cruzi* foi pouco estudada até o momento. Por isso, o mapeamento de suas funções se torna crucial (MOHAMED; SLOANE, 2006). O uso de ferramentas biológicas como o CRISPR/Cas9, que revolucionou a manipulação genética, e pode ser aplicada para estudos em organismos que possuem alta complexidade molecular e representam um desafio, como o *T. cruzi*, tem potencial de facilitar a manipulação do genoma do parasito. Consequentemente, também no entendimento das complexidades envolvendo interações patógeno-hospedeiro (PENG et al., 2015; LANDER; CHIURILLO; DOCAMPO, 2016).

Além disso, conhecimento da estrutura tridimensional desta enzima pode promover um melhor entendimento de suas funções e formas de inibir sua atividade (JUMPER et al., 2021). O uso de estratégias computacionais, permite a elucidação de estrutura 3D dos alvos e seleção de potenciais inibidores desses alvos, otimizando tempo, recursos e equipamentos experimentais para priorização de compostos.

# **OBJETIVOS**

### 3 OBJETIVOS

#### 3.1 Objetivo geral

Este trabalho tem como objetivo realizar um estudo estrutural da catepsina B de *Trypanosoma cruzi* por meio de abordagens *in silico* e *in vitro*.

#### 3.2 Objetivos específicos

1. Desenhar e sintetizar o sgRNA para *cattc*;
2. Construir um cassete de homologia para o gene *catbtc* contendo um gene de resistência blasticidina S deaminase;
3. Transfectar os parasitos das linhagens *CL Brener* e *Y* com o sgRNA e o cassete de homologia e em seguida, selecionar os mutantes editados;
4. Confirmar a edição dos parasitos clonados por PCR e *Western Blot*;
5. Expressão heteróloga da catepsina B de *Trypanosoma cruzi* recombinante em *E. coli*.
6. Gerar e avaliar modelos da estrutura tridimensional da catepsina B de *Trypanosoma cruzi* através de metodologias computacionais;
7. Realizar a triagem virtual baseada em *docking* e em modelos de QSAR fenotípicos para priorização de compostos inibidores da catepsina B de *Trypanosoma cruzi*.

# **METODOLOGIA**

## 4 METODOLOGIA

### 4.1 Metodologia computacional

#### 4.1.1 Análise in silico da catepsina B de *Trypanosoma cruzi*

Os 2 alelos provenientes da cepa CL Brener do gene *catbtc*, assim como a sequência de aminoácidos que correspondem a proteína expressa por cada alelo foram obtidos na plataforma TriTrypDB (alelo *Esmeraldo-like* <https://tritrypdb.org/tritrypdb/app/record/gene/TcCLB.510535.100> e alelo *Non-Esmeraldo-like* <https://tritrypdb.org/tritrypdb/app/record/gene/TcCLB.511827.100>). Os alinhamentos destes alelos e da sequência da proteína foram realizados pelos programas MUSCLE (ClustalW) <https://www.ebi.ac.uk/Tools/msa/muscle/> e BLAST <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Análises *in silico* da CatBTc foi realizada para explorar características físico-químicas da proteína expressa e a sequência de aminoácidos correspondente ao alelo *Esmeraldo-like* foi também observada. Os servidores utilizados foram: OrthoDB (ZDOBNOV et al., 2021), que cataloga hierarquicamente os ortólogos de um gene; IQ-TREE Web Server (TRIFINOPOULOS et al., 2016), com o objetivo de montar uma árvore filogenética de proteínas relacionadas à catepsina B de *T. cruzi* obtidas no BLAST; ProtParam (HOCHSTRASSER, 2005), o qual fornece os parâmetros de massa molecular, ponto isoelétrico teórico e índice de instabilidade; SignalP 5.0 (ALMAGRO ARMENTEROS et al., 2019), que prediz a presença ou ausência, assim como a localização de sítios de clivagem de peptídeo sinal; TMHMM 2.0 (MÖLLER; CRONING; APWEILER, 2001), que indica possíveis domínios transmembrana da proteína; os servidores NetNGlyc 1.0 (GUPTA; BRUNAK, 2002) e NetOGlyc 4.0 (STEENTOFT et al., 2013), os quais predizem sítios de N-glicosilação e O-glicosilação, respectivamente; o servidor ESPript 3.0 (ROBERT; GOUET, 2014), que possibilita o alinhamento completo de sequências primárias e as compara com uma estrutura secundária proveniente de um PDB selecionado; o servidor NetSurfP-2.0 (KLAUSEN et al., 2018), que prediz a estrutura secundária de uma proteína a partir de sua sequência primária; a plataforma STRING (SZKLARCZYK et al., 2019) prediz interações proteína-proteína descritas na literatura, incluindo associações diretas e indiretas.

#### 4.1.2 Modelagem da estrutura tridimensional da catepsina B de *Trypanosoma cruzi*

##### 4.1.2.1 Predição da estrutura tridimensional

A estrutura tridimensional cristalizada da catepsina B de *Trypanosoma cruzi* ainda não foi obtida experimentalmente. Por isso, realizamos a predição do modelo tridimensional dessa estrutura por meio de quatro servidores SwissModel, Itasser, trRosetta e AlphaFold, que utilizam diferentes métodos: homologia, enovelamento e aprendizado de máquina, respectivamente.

A sequência primária completa da catepsina B foi extraída da plataforma Uniprot (<https://www.uniprot.org/uniprot/Q4DE83>), pelo código Uniprot Q4DE83. A sequência primária foi submetida no servidor Swiss Model (SCHWEDE et al., 2003) e as 2 estruturas já resolvidas (PDB 4HWY (REDECKE et al., 2013) e PDB 3HHI (KERR et al., 2010)) com maior identidade sequencial foram selecionados para gerar modelos. Dos parâmetros observados no Swiss model, a identidade sequencial, cobertura e o QMEAN foram considerados. O QMEAN qualifica em valores potenciais estatísticos dos campos de força das ligações das proteínas e a consistência do modelo com características as estruturais provenientes da sequência, o qual varia geralmente de 0 a 1 (STUDER et al., 2020).

A sequência primária da proteína foi submetida posteriormente no servidor I-Tasser (YANG; ZHANG, 2015). Para gerar os modelos, utilizamos o modo padrão e a opção de modelar a partir de um molde de PDB específico. Dessa forma, a modelagem a partir dos PDB 4HWY e 3HHI foi efetuada. O parâmetro considerado pelo I-Tasser para verificar a qualidade dos modelos é o C-score, o qual atribui uma pontuação de confiabilidade ao modelo gerado que varia de -5 a 2, em que os melhores modelos se aproximam de 2, conferindo assim uma qualidade global estimada ao modelo baseada computacionalmente na qualidade do *template*, assim como na convergência da simulação REMC (Replica-Exchange Monte Carlo) (ZHENG et al., 2019). Outro parâmetro analisado foram os PDB moldes buscados pelo servidor e utilizados na modelagem, visto que este programa prepara os modelos em base do enovelamento de outros PDB já existentes e os lista na ordem dos 10 melhores.

O modelo gerado no servidor trRosetta (YANG et al., 2020) foi realizado no modo padrão. Para o trRosetta, o parâmetro de qualidade é o TM-score que mostra a estimativa da confiabilidade do modelo, o qual se baseia na probabilidade da distância e convergência do melhor modelo com as dos demais modelos previstos, o qual varia de 0 a 1. A partir do valor de 0,6 de TM-score a topologia do modelo foi corretamente prevista.

O modelo AlphaFold foi originado no modo padrão a partir do código Uniprot Q4DE83, que diz respeito à catepsina B de *T. cruzi* da cepa CL Brener alelo *Esmeraldo-Like*. Como parâmetro principal, o programa utiliza a métrica pLDDT para prever as diferenças de distância local de cada resíduo da proteína fornecendo assim um score de confiabilidade com pontuação entre 0 e 100. Essa pontuação é dividida em 4 subcategorias: muito baixo com scores menores que 50; baixo, entre 50 e 70; confiável, entre 70 e 90 e muito alto, acima de 90. Cada resíduo da proteína recebe uma subcategoria que é separada por cor para realçar a qualidade do modelo final construído (JUMPER et al., 2021).

#### 4.1.2.2 Validação da qualidade geométrica e estereoquímica

A qualidade dos modelos gerados foi analisada através do servidor *Molprobability* (WILLIAMS et al., 2018), no qual foram examinados os parâmetros: *Ramachandran favored*, *Clashscore*, e *Molprobability score*. Os gráficos de Ramachandran relatam a distribuição dos ângulos de torção ( $\phi$  e  $\psi$ ) da cadeia principal da proteína e são uma das métricas de qualidade mais utilizadas para avaliar modelos de estrutura experimental. O parâmetro *Ramachandran favored*, que indica, dentro do gráfico de Ramachandran, a porcentagem de resíduos pertencentes a regiões favorecidas, mais estáveis dos aminoácidos. Esse parâmetro, portanto, faz a análise de cada resíduo em torno dos seus ângulos de torção e os compara com a distribuição esperada deles, mostrando se estão em regiões favoráveis ou não.

O parâmetro *clashscore* indica se há sobreposição entre átomos (choques estéricos) maior ou igual a 0,4 Å a cada 1000 átomos na estrutura da proteína.

Por fim, o parâmetro *Molprobability Score* é uma combinação dos parâmetros *clashscore*, *Ramachandran not favored* e *bad side-chain rotamers* em um único número, normalizado para estar na mesma escala da resolução de raio-X, refletindo

a resolução cristalográfica de uma proteína. Os valores de *Molprobability Score* e *clashscore* são comparados através de um percentil relativo a estruturas próximas da mesma resolução obtida, concedendo assim, poder de comparação das entradas no programa com diferentes resoluções encontradas para sua molécula de interesse. Os valores desejáveis desses parâmetros são: *Molprobability Score* com percentil maior ou igual a 66%, *Ramachandran favored* com o favorecimento maior ou igual a 98% e *clashscore* com percentil maior ou igual a 66. O servidor GalaxyRefine (KO et al., 2012) foi utilizado para o refinar os modelos gerados. Os resultados obtidos foram reavaliados pelo *Molprobability*, em alguns casos, os modelos foram refinados mais de uma vez.

#### 4.1.3 Busca por compostos para validação do protocolo de *docking*

Em paralelo à modelagem, uma busca por inibidores da catepsina B de *T. cruzi* foi realizada. Como não existem inibidores específicos para a CatBTc, esta busca se estendeu a inibidores conhecidos de catepsinas B de *H. sapiens* e organismos próximos de *T. cruzi*, como outros tripanossomatídeos. Através da pesquisa, foram obtidos na plataforma ChEMBL (ChEMBL5832), um compilado de resultados de 3 artigos sobre o *design* de 88 compostos testados *in vitro* para a inibição de catepsina B de *Trypanosma brucei* (TbCatB), os quais serviram de parâmetro para a escolha de compostos inibitórios para *T. cruzi* e para o *docking* pelos programas. Como um limiar de 500nM de IC50 havia sido selecionado, uma lista própria de 14 compostos (**Figura 11**) extraída de 5 artigos diferentes (CIANNI et al., 2021; JÍLKOVÁ et al., 2021; MALLARI et al., 2009; NDAO et al., 2014; SCHMITZ et al., 2016) foi elaborada em base de compostos altamente inibitórios e seletivos da catepsina B de diferentes organismos. O resultado foi os compostos considerados ativos usados neste trabalho. Um conjunto de compostos *decoys* foi construído através do *workspace* Knime (BERTHOLD et al., 2006) a partir da dataset ZINC-15 (STERLING; IRWIN, 2015) em que compostos inativos para o alvo foram localizados em base da similaridade estrutural com os compostos considerados ativo, sendo assim, aplicados às métricas em relação aos resultados dos *dockings*. Este conjunto formado por 504 *decoys*, com similaridade estrutural de 0,3 a 0,75 em relação aos ativos, seguiu a proporção de 36 compostos *decoys* para cada composto ativo, totalizando 518 compostos.

#### 4.1.4 Preparo dos ligantes e proteína

Os compostos passaram por um filtro de remoção de duplicatas, utilizando o programa Knime (FILLBRUNN et al., 2017). Todos os ligantes foram preparados utilizando a ferramenta LigPrep (SCHRÖDINGER RELEASE, 2021) do Glide nos pHs de  $4,5 \pm 0,5$  e  $7,4 \pm 0,5$ , corrigindo o estado de protonação de acordo com a ferramenta Epik (GREENWOOD et al., 2010), gerando até 10 confôrmeros por ligante e mantendo a quiralidade dos compostos quirais. As coordenadas do grid (eixo X: 5.602; eixo Y: -0.347; eixo Z: -3.427) das proteínas foram baseadas tríade catalítica da catepsina B, que corresponde ao mesmo sítio do ligante co-cristalizado de estrutura cristalográfica da proteína homóloga PDB 3HHI.

Para o programa DockThor, foi selecionada a opção *Virtual screening*, e foram mantidos no valor padrão do programa: ajustes de número de avaliações, tamanho da população, *seed* inicial e número de corridas. O programa Glide foi utilizado no modo extra precisão (XP) (FRIESNER et al., 2006).

#### 4.1.5 *Docking* molecular

Os programas DockThor VS web (DE MAGALHÃES; BARBOSA; DARDENNE, 2004b, 2004a; SANTOS et al., 2020) e Glide versão XP (*Grid-Based Ligand Docking with Energetics*), que faz parte do pacote do *workspace* Maestro (FRIESNER et al., 2006) foram utilizados para realização dos *dockings* dos modelos de catepsina B de *T. cruzi*.

As proteínas foram preparadas no programa Glide usando a ferramenta Protein Preparation Wizard (MADHAVI SASTRY et al., 2013; SCHRÖDINGER, 2015) em pHs de  $4,5 \pm 0,5$  (pH ótimo de atividade da catepsina B) e  $7,4 \pm 0,5$  (pH de atividade de endopeptidase), seguindo o protocolo de adição de átomos de hidrogênio e minimização de energia do sistema, utilizando campo de força OPLS-2005 (HAYES; STEIN; WEISER, 2004)

#### 4.1.6 Validação do protocolo de *docking* molecular

A validação do protocolo de *docking* foi realizada através de compostos ativos e *decoys*, utilizando o protocolo de *docking* previamente citado. Primeiramente, os compostos foram preparados e submetidos ao *docking* nos dois programas DockThor e Glide, nos pHs de 4,5 e 7,4 e passaram por avaliação das métricas obtidas. Os pHs diferentes foram selecionados considerando a possibilidade de as estruturas tridimensionais estarem em conformações distintas devido à dupla atividade de endo e exopeptidase.

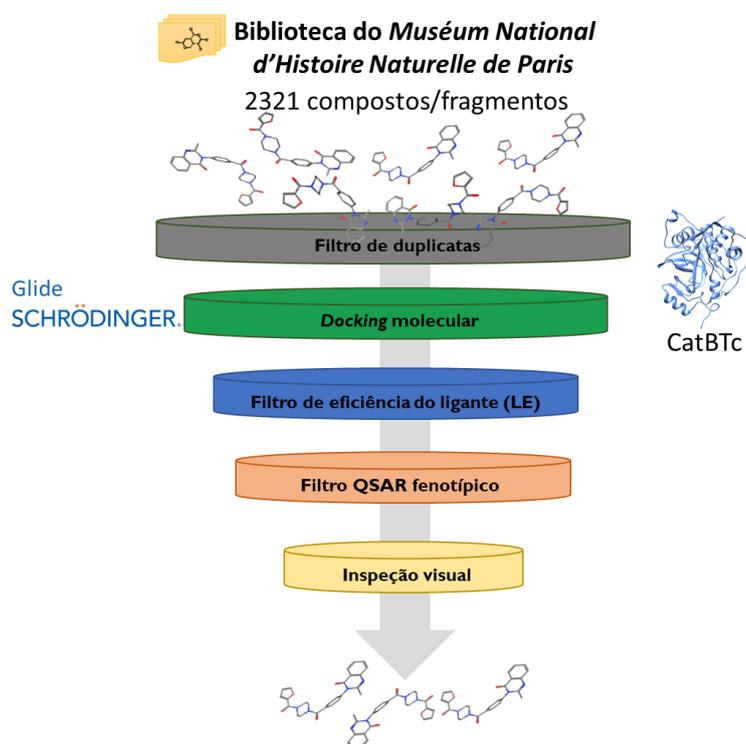
Os compostos foram divididos em populações de *top 1%*, *top 5%* e *top 10%* de acordo com os *scores* mais negativos alcançados. As seguintes métricas foram avaliadas durante essa etapa: (i) sensibilidade (Se), sendo esta o número de verdadeiros positivos dividido pela soma de verdadeiros positivos e falso negativos; (ii) especificidade (Sp), indicando o número de verdadeiros negativos dividido pela soma de verdadeiros negativos e falso positivos; (iii) CCR ( taxa de classificação correta), que é a média da soma da sensibilidade e especificidade; (iv) área sob a curva ROC (AUC), onde é formada uma curva ROC em um gráfico pela variação de limiares da sensibilidade, no eixo y e da taxa de falsos positivos (1- especificidade), no eixo x. A área designa a classificação global do modelo, que quanto mais próximo de 1, melhor; (v) fator de enriquecimento (EF) especifica quão melhor é o modelo em ordenar os compostos comparado a uma lista aleatória; (vi) BEDROC (Discriminação aprimorada de Boltzmann da curva ROC) que mostra a probabilidade dos compostos ativos serem previstos no topo da lista comparado a uma lista aleatória (MELO-FILHO, 2018).

Essa validação se torna importante no momento da escolha, através dos parâmetros anteriormente citados, do melhor protocolo de *docking* a ser utilizado na etapa de triagem virtual.

#### 4.1.7 Triagem virtual

Realizamos uma triagem virtual da biblioteca do Muséum National d'Histoire Naturelle de Paris (Museu Nacional de História Natural de Paris - MNHN) através de diferentes abordagens computacionais, incluindo *docking* molecular contra catepsina B de *T. cruzi*, além dos filtros de eficiência de ligante, similaridade estrutural e QSAR fenotípico. Essa biblioteca do MNHN já foi bastante explorada pelo nosso grupo LIPH

na busca de compostos reativos a diferentes proteases estudadas no laboratório. Um fluxograma geral das abordagens aplicadas está representado na **Figura 12**.



**Figura 12.** Esquema das etapas da triagem virtual dos compostos provenientes do Museu Nacional de História Natural de Paris (MNHN). As etapas seguem a seguinte sequência: filtro de duplicatas; *docking* molecular dos compostos contra o modelo 3D da catepsina B de *T. cruzi*; filtro de eficiência de ligante (LE); filtro fenotípico de QSAR e inspeção visual.

#### 4.1.8 Filtro de eficiência de ligante

Os *top* 10% compostos resultantes do *docking* foram filtrados pelo parâmetro eficiência do ligante (LE), que relaciona a energia livre de ligação ou *score* de *docking* ( $\text{Kcal} \cdot \text{mol}^{-1}$ ) ao peso molecular dos átomos pesados do composto. Valores de LE acima de  $0,3 \text{ Kcal} \cdot \text{mol}^{-1} \cdot \text{átomo não-hidrogênio}^{-1}$  são considerados aceitáveis para fármacos. Tal parâmetro normaliza compostos pequenos e grandes, isso permite verificar a eficiência de um ligante e reduzir a diferença entre compostos com estruturas pequenas em relação às estruturas grandes, que fazem mais interações nos cálculos de *docking*. Além disso, possibilita uma melhor avaliação da potência de ligação em seu nível molar, não apenas se baseando no  $\text{IC}_{50}$  ou  $\text{K}_i$  (POLANSKI; TKOCZ; KUCIA, 2017).

#### 4.1.9 Filtro de QSAR fenotípico

Os compostos foram então submetidos a um segundo filtro de QSAR (*Quantitative Structure-Activity Relationships*) fenotípico de *T. cruzi*, no intuito de classificar os compostos em ativos e inativos frente ao parasito. Os modelos de QSAR fenotípicos utilizados no trabalho foram desenvolvidos no laboratório do Prof. Dr. Bruno Junior Neves, LabMol (Universidade Federal de Goiás). Os modelos foram empregados como filtros para predição de atividade em *T. cruzi* dos compostos e o critério de filtragem foi a probabilidade de ser ativo igual ou superior a 0,6.

Os modelos de classificação foram construídos através de aprendizado de máquina, através de aprendizado profundo (redes neurais), utilizando dados experimentais para *T. cruzi* com concentração inibitória de 50% (IC<sub>50</sub>) reportada, em ensaios de 72h e 96h, extraídos da base de dados ChEMBL (GAULTON et al., 2012). Um limiar de atividade de 10 µM foi utilizado para classificar os compostos em ativos (IC<sub>50</sub> ≤ 10 µM) e inativos (IC<sub>50</sub> > 10 µM) (KATSUNO et al., 2015). O modelo foi desenvolvido seguindo as melhores práticas de modelagem preditiva e princípios da OECD para construção e validação de modelos de QSAR.

#### 4.1.10 Filtro de similaridade estrutural e agrupamento e compostos

Paralelamente a este filtro, o programa DataWarrior (SANDER et al., 2015) foi empregado para análise de *clusters*. Essa análise foi baseada em agrupamento por similaridade estrutural de >80% entre os compostos. Para o cálculo de similaridade, é necessária a utilização de um descritor molecular, o qual tem como função transformar a informação química codificada representada simbolicamente por uma molécula em um número útil através de um procedimento lógico e matemático (ALVES et al., 2018). O descritor utilizado foi o SkelSpheres que é aplicado quando se exige uma análise estrutural mais refinada, pois considera a estereoquímica, os heteroátomos e possui uma melhor resolução propiciando menos colisões graves. (SANDER et al., 2015).

#### 4.1.11 Inspeção visual dos compostos

Por fim, a inspeção visual das poses de *docking* dos compostos foi realizada para analisar se as interações ocorriam com resíduos importantes (sítio catalítico e

histidinas no *loop* de oclusão), além de sobreposição com estrutura de diferentes PDBs (3HHI; 1GMY (GREENSPAN et al., 2001); 6AY2 (WEI et al., 2018); 3AI8 (MIRKOVIĆ et al., 2011)) contendo ligantes co-cristalizados. As interações foram inspecionadas com o servidor ProteinsPlus (SCHÖNING-STIERAND et al., 2020), com a ferramenta PoseView (STIERAND; MAASS; RAREY, 2006) que gera imagens bidimensionais das interações proteína-ligante nas poses fornecidas.

## 4.2 Metodologia experimental

### 4.2.1 Cultura de células

As formas epimastigotas selvagens de *T. cruzi* das cepas CL Brener e Y foram cultivadas em meio *Liver Infusion Tryptone* (LIT) suplementado com 10% (v/v) soro fetal bovino (SFB) e penicilina/streptomomicina (1x) à 28 °C. Os parasitos transfectados com o pTREX-n/ Cas 9 e o cassete foram cultivados nas mesmas condições dos selvagens com o acréscimo de G418 (150µg/ml) e blasticidina (25µg/ml).

### 4.2.2 Desenho do sgRNA para nocaute com CRISPR/Cas9

Com o objetivo de causar o nocaute do gene *catbtc*, gerando assim, parasitos mutantes nulos por meio da utilização do sistema CRISPR/Cas9, houve o desenho de um sgRNA como primeiro passo. A ferramenta EuPaGDT (<http://grna.ctegd.uga.edu>) foi usada com fim de gerar uma lista de sgRNA em que foi escolhido o melhor de acordo com o objetivo de guiar o corte da endonuclease Cas 9 no gene *catbtc* baseado na conservação dos dois alelos presentes TcCLB.510535.100 (*Esmeraldo-Like*) e TcCLB.511827.100 (*Non-Esmeraldo-Like*), na cepa CL Brener, depositados no TriTrypDB (<https://tritrypdb.org/tritrypdb/>). Dentre os critérios utilizados para avaliar o melhor sgRNA estavam o *score* obtido pela plataforma, a posição do corte em relação ao gene completo, a ausência de *off-targets* e a não geração de estruturas secundárias de grampo. Assim, foi possível obter um *primer* senso denominado *CatTcB331\_Fwd* e um *primer* antisenso denominado *sgRNA-scaffold-Common-rvs* com capacidade de amplificar o *backbone* do sgRNA de 82 pb do plasmídeo pUC\_sgRNA que foi posteriormente utilizado para a amplificação do sgRNA, resultando no sgrNA final contendo 122 pb. Além disso, os *primers* possuem sítio de

restrição para *Bam*HI possibilitando a futura clonagem no plasmídeo pTREX-n/ Cas 9 (Tabela 1).

**Tabela 1 - Primers utilizados na metodologia CRISPR/Cas9**

<b>Primers</b>	<b>Sequência</b>
<i>CatTcB331_Fwd</i>	5'-GATCGGATCCGATCCGTGACCAATCGAGCTGGTTTTAGAGCTAGAAATAGC-3'
<i>sgRNA-scaffold-Common-rvs</i>	5'-CAGTGGATCCAAAAAGCACCGACTCGGTG-3'
<i>HX1_rvs</i>	5'-TAATTCGCTTTCGTGCGTG-3'
<i>sgRNA final</i>	5'-GATCGGATCCGATCCGTGACCAATCGAGCTGGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT TTTTGGATCCACT-3'

Destacado de azul está o sítio de restrição de *Bam*HI e destacado de amarelo, a sequência de 20 nucleotídeos proveniente do gene *catbtc* para o corte do CRISPR/Cas9.

#### 4.2.3 Amplificação e purificação do sgRNA

A amplificação do sgRNA foi realizada por PCR usando 25ng do plasmídeo pUC\_sgRNA, 300 nM de cada *primer* e a Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific™) para uma reação final de 50 µl. O programa seguido foi: 95 °C de desnaturação inicial por 2 minutos, 95 °C de desnaturação por 30 segundos, 58 °C para o anelamento por 30 segundos, 72 °C de extensão por 1 minuto, 72 °C de extensão final por 5 minutos em 30 ciclos de repetição. O produto de PCR produzido foi confirmado por gel de agarose (2,5%). O sgRNA foi purificado utilizando QIAquick® Gel Extraction Kit (Qiagen™) e eluído em 50 µl. E posteriormente, quantificado pelo NanoDrop™ Lite Spectrophotometer (Thermo Scientific™). Após, foi realizada uma digestão com a enzima de restrição *Bam*HI (Jena Bioscience™) para a clonagem no plasmídeo pTREX-n/ Cas 9.

#### 4.2.4 Preparação do plasmídeo pTREX-n/Cas9 e ligação ao sgRNA

O plasmídeo pTREX-n/Cas9 foi preparado através de Midi-prep utilizando PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen™), quantificado pelo NanoDrop™ Lite Spectrophotometer (Thermo Scientific™) e linearizado com a enzima de restrição *Bam*HI (Jena Bioscience™). Em seguida, o pTREX-n/Cas9 foi desfosforilado com a

enzima Shrimp Alkaline Phosphatase (SAP) (Promega™) por 1 hora a 37 °C e inativada por 20 min a 65 °C. O plasmídeo, então, foi purificado por gel de agarose 0,8% (m/v) utilizando QIAquick® Gel Extraction Kit (Qiagen™) e eluído em 40 µl. A ligação foi realizada com 120 ng do vetor pTREX-n/Cas9 na proporção 5:1 com o inserto sgRNA com a ação da enzima T4 DNA ligase (Promega™) por 16 horas a 4 °C para um total final de 13 µl de reação.

#### 4.2.5 Transformação em bactérias competentes

O plasmídeo pTREX-n/Cas9/sgRNA resultante da ligação foi transformado em bactérias *E. coli* cepa TOP10 previamente manipuladas para se tornarem competentes e serem capazes de integrar DNA exógeno. Foram adicionados à bactéria 2 µl da ligação e deixados no gelo por 30 minutos e, a seguir, um choque térmico de 42 °C por 2 minutos, logo seguido de 2 minutos no gelo. Então, foi acrescentado 900 µl de meio SOC e cultivado no *shaker* com rotação de 220 rpm a 37 °C por 1 hora. Assim, foram plaqueados em meio LB ágar contendo 100 µg/ml de ampicilina a 37 °C por 16 horas.

#### 4.2.6 Confirmação dos clones transformados de pTREX-n/Cas9/sgRNA

Foram selecionadas aleatoriamente 21 colônias crescidas nas placas. Essas colônias foram denominadas colônias 1 a 21 e foram plaqueadas novamente em meio LB ágar. Essas colônias foram coletadas e submetidas a uma PCR para confirmação da clonagem. O DNA das colônias foi extraído pela suspensão em 10µl de água ultrapura estéril e pasteurização em 90 °C por 10 minutos, além de 600 nM de cada *primer*, 200 mM dNTP, 1,25 U de *Taq* DNA Polymerase Recombinant (Invitrogen™), 1,5 mM MgCl<sub>2</sub>, e tampão 10X para reação final de 25 µl. Os *primers* usados nessa reação foram: *CatTcB331\_Fwd* como *primer* senso e *HX1\_rvs* como *primer* antisenso (**Tabela 1**). Os parâmetros da PCR foram as seguintes: desnaturação inicial a 94 °C por 3 minutos, desnaturação a 94 °C durante 45 segundos, anelamento a 55 °C por 30 segundos e extensão na temperatura de 72 °C por 30 segundos, repetindo por 35 ciclos, com extensão final a 72 °C durante 10 min. A amplificação foi confirmada por gel de agarose 2,5 % (m/v). Apenas dois clones foram selecionados e pré-inoculados

em 5 ml de meio LB contendo 100 µg/ml de ampicilina e crescidos em 37 °C por 3 horas e, em seguida, inoculado em 100 ml de meio LB com concentração igual de ampicilina e crescido nas mesmas condições do pré-inóculo. Então, foi preparado o DNA plasmidial através de midi-prep com PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen™) e posteriormente quantificados com NanoDrop™ Lite Spectrophotometer (Thermo Scientific™).

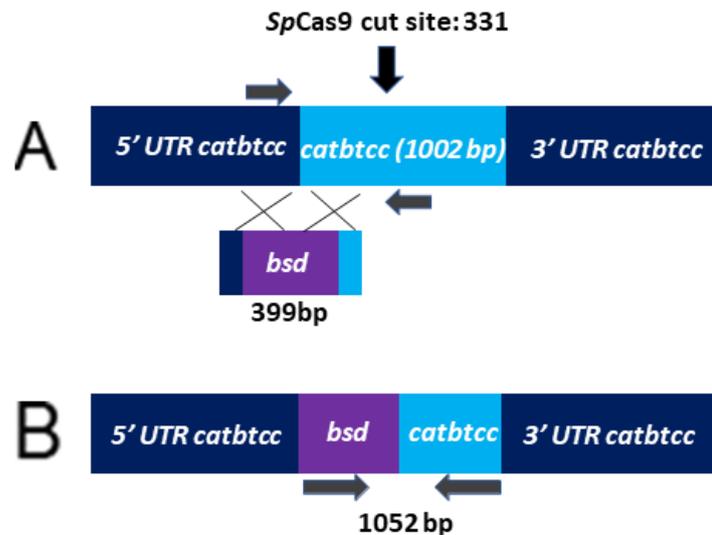
#### 4.2.7 Produção de cassete para reparo por homologia (HDR)

Além do plasmídeo pTREX-n/Cas9/sgRNA, um cassete foi produzido para guiar o reparo por homologia. O primeiro passo para a produção foi o desenho de *primers* para realizar a amplificação por PCR do cassete. O desenvolvimento foi feito por meio da sequência de DNA do alelo *Esmeraldo-Like* da catepsina B de *T. cruzi*. Foram selecionados *primers* tanto para a amplificação do cassete, *CatBTc Ult. Forward*, como senso e *CatBTc Ult. Reverse*, como anti-senso, assim como *primers* para confirmação da inserção do cassete, *CatBTc Conf. Forward*, como senso e *CatBTc Conf. Reverse*, como anti-senso (**Tabela 2**). O *primer* de amplificação senso foi desenhado com os 80 pb antecedentes ao ATG do gene da catepsina B, ou seja, na região 5' UTR mais os 20 primeiros pb do gene da blasticidina. O *primer* de amplificação anti-senso possui os 20 últimos pb do gene da blasticidina mais os 80 pb da região à jusante ao corte feito pelo sgRNA, assim resultando em um cassete de 556 pb contendo o gene completo da blasticidina, que funciona como a marca de seleção do cassete, intercalado pela região 5' UTR do gene da catepsina B e a região adjacente ao corte feito pela endonuclease Ca9. Assim, resultando no gene da catepsina B (*catbtc*) truncado com a inserção do gene bsd com o tamanho final de 1052 pb (**Figura 13**). Foram desenhados também *primers* que amplificam uma região de 181 pb dentro do gene da blasticidina (*bsd*), vistos na **Tabela 2**.

**Tabela 2 - Primers** usados na metodologia de amplificação e confirmação do cassete de homologia

<b>Primers</b>	<b>Sequência</b>
<i>CatBTc Ult. Forward</i>	5'-AAAAGGGTTACAACCTGCTTAAAAGGAAGAGACGAAAAACAACAACAACA AAGTTACGGGTGGAAGAAATCAGAGATGGCCAAGCCTTTGTCTCAAG-3'
<i>CatBTc Ult. Reverse</i>	5'-GGTTATGTGTGGGAGGGCTAACGGCTCTTGCTGGGCTGTGGCAGCAGCTTCA GCCATTCGGACCGCTATTGCACTCTTGGTGGCGTACGGGACCTCAGA -3'
<i>bsd Forward</i>	5'- CCATCTCTGAAGACTACAGC -3'
<i>bsd Reverse</i>	5'- GCGATCGGAAATGAGAACAG -3'

***CatBTc Ult. Forward* e *CatBTc Ult. Reverse* contendo 100 pb cada com regiões destacadas de azul indicando os primeiros e últimos 20 pb do gene *bsd*, respectivamente. *CatBTc Ult. Forward* contém 80 pb referentes à região 5' UTR, anteriores ao ATG do gene *catbtcc* e *CatBTc Ult. Reverse* possui 80 pb da região posterior ao corte da *SpCas9* no gene *catbtcc*, ambos destacados de amarelo. *bsd Forward* e *bsd Reverse* com primers que amplificam dentro do gene *bsd*.**



**Figura 13. Esquema do reparo por homologia causado pelos ultrâmeros no local de corte da *SpCas9*.** **A.** Gene *catbtcc* contendo 1002 pb de azul claro com a presença das regiões 5'UTR e 3'UTR do gene de azul escuro. Seta vertical indicando o sítio de corte da *SpCas9* na posição 331 dentro do gene *catbtcc*. Gene da blasticidina (*bsd*) contendo 399 pb de roxo com a inclusão dos ultrâmeros provenientes da região 5'UTR *catbtcc* (azul escuro) e após o sítio 331 dentro do gene *catbtcc* (azul claro) indicando a região de reparo por homologia direta realizada após o corte da *SpCas9*. **B.** Gene da blasticidina incorporado dentro do gene da *catbtcc* contendo 1052 pb.

#### 4.2.8 Amplificação do cassete de reparo por homologia (HDR)

A PCR para amplificação do cassete foi realizada com o uso de 400 ng de cada *primer*, 1,6 ng de DNA proveniente do plasmídeo pGEM-BSD contendo o gene da blasticidina previamente clonado e 12,5 µl da enzima 2x Phusion Flash High-Fidelity

PCR Master Mix (Thermo Scientific™) para uma reação final de 25 µl. Os parâmetros utilizados foram: desnaturação inicial 93 °C por 3 minutos, desnaturação 94 °C por 45 segundos, anelamento 60 °C por 45 segundos, extensão 72 °C por 1 minuto e extensão final 72 °C por 10 minutos com 34 ciclos de repetição. A confirmação foi realizada por gel de agarose 1,5% (m/v).

#### 4.2.9 Clonagem e transformação do cassete amplificado

Após a confirmação da amplificação, o produto de PCR foi adicionado de 1 µl da enzima *Taq* DNA Polymerase Recombinant (Invitrogen™) e incubada 72 °C por 10 minutos para causar uma adição de uma cauda poli A nas extremidades do cassete, facilitando assim, a posterior clonagem no vetor pGEM®-T Easy (Promega™) (**Figura 14**). Então, o cassete foi acrescentado de T4 DNA ligase (Promega™) e o vetor pGEM para inserção para uma reação final de 10 µl, a 4 °C por 16 horas. Depois, 4 µl da ligação foi inoculada em bactérias competentes *E. coli* cepa TOP 10 para realizar a transformação desse plasmídeo seguindo o protocolo descrito anteriormente. A diferença foi no plaqueamento em que as placas de LB ágar foram incluídas de IPTG e X-Gal para a seleção das colônias ser feita através do processo de *White/Blue* com a interrupção do gene da β-galactosidase (LANGLEY et al., 1975; ULLMANN; JACOB; MONOD, 1967).

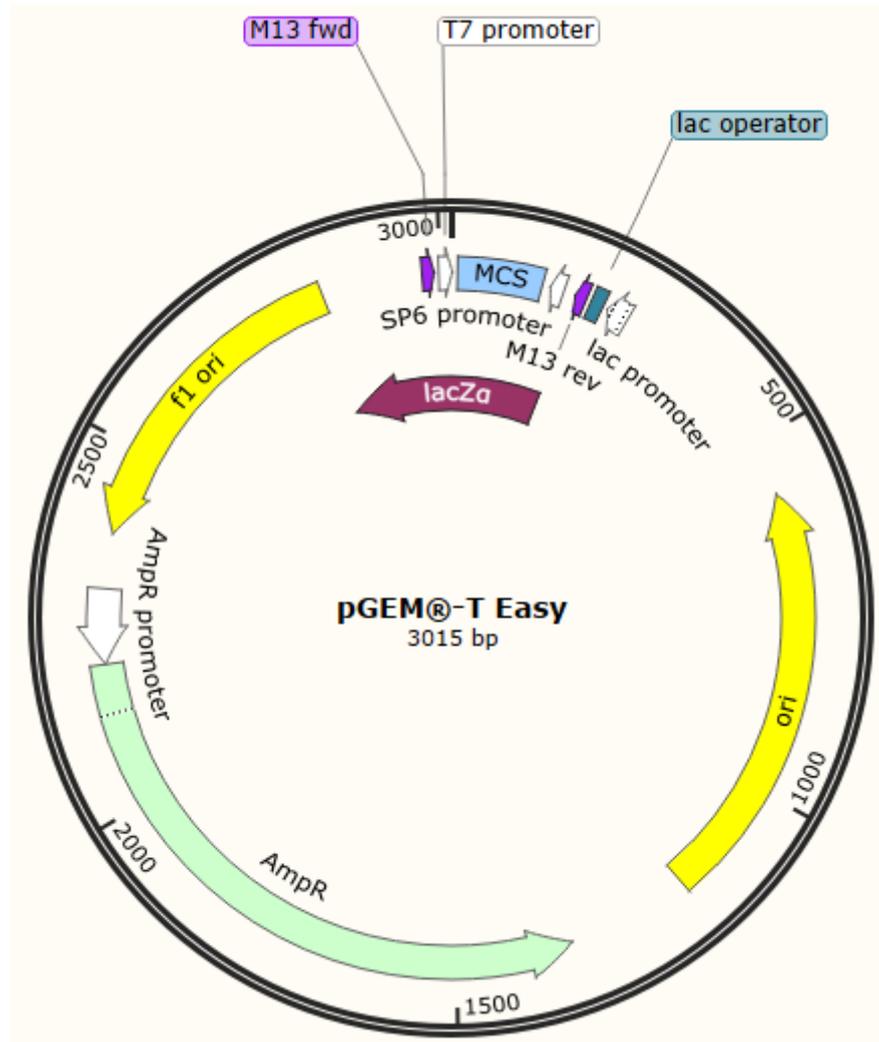


Figura 14. Representação do vetor pGEM®-T Easy (Promega™) utilizado na clonagem do cassete visto no programa SnapGene importado das sequências online do próprio programa. Representação dos componentes desse vetor incluindo o sítio de múltipla clonagem.

#### 4.2.10 PCR de colônia dos transformados com o cassete

A confirmação da clonagem ocorreu por PCR. As colônias foram escolhidas da seguinte forma: 3 colônias brancas, que teoricamente seriam positivas e 1 colônia azul, que teoricamente seria negativa, sendo todas elas escolhidas aleatoriamente. O DNA foi lisado da mesma forma que foi descrito anteriormente no tópico 4.7 (Confirmação dos clones transformados de pTREX-n/Cas9/sgRNA). Os parâmetros utilizados na PCR foram os seguintes: desnaturação inicial 93 °C por 3 minutos, desnaturação 94 °C por 45 segundos, anelamento 55 °C por 45 segundos, extensão 72 °C por 30 segundos e extensão final 72 °C por 10 minutos com 34 ciclos de repetição. A confirmação se deu por gel de agarose 2% (m/v) com uma amplificação

de 181 pb. Das colônias positivas, o clone 3 foi selecionado e pré-inoculado em 2 ml de meio LB com 100 µg/ml de ampicilina a 37°C por 3 horas e, então inoculado em 100 ml de meio LB com a mesma concentração de ampicilina a 37 °C por 16 horas. Após, a preparação de DNA plasmidial foi realizada com PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen™) e quantificados com NanoDrop™ Lite Spectrophotometer (Thermo Scientific™).

#### 4.2.11 Transfecção do plasmídeo pTREX-n/Cas9/sgRNA em *T. cruzi* CL-Brener e Y

O plasmídeo pTREX-n/Cas9/sgRNA foi pasteurizado 65 °C por 30 minutos. Culturas de epimastigotas das cepas CL Brener e Y selvagens contendo  $1 \times 10^6$  parasitos por tubo foram lavadas com 1 ml de PBS com rotação de 2000 g por 15 minutos, o sobrenadante foi retirado e ressuscendido novamente em 1 ml de PBS para uma nova lavagem nas mesmas condições anteriores e a retirada de todo o sobrenadante. O aparelho utilizado para a transfecção foi o Amaxa® Nucleofector™ 2b Device (Lonza™) que disponibiliza no seu kit Human T Cell Nucleofector™ o tampão de ressuspensão e as cubetas para a transfecção. Para que o tampão possa ser utilizado, é necessária a mistura de 2 reagentes que vêm juntamente ao kit, então, 82 µl de Nucleofector™ Solution for Human T Cells foi misturado com 18 µl de Supplement 1 (Lonza™) resultando em 100 µl de tampão para cada amostra. Assim, o pellet de parasitos foi ressuscendido no tampão. Foram feitas duas tranfecções para cada cepa: uma apenas com a adição de PBS para servir de controle e outra com adição de 25 µg do pTREX-n/Cas9/sgRNA. Após a transfecção, as células ficaram incubadas 5 minutos no gelo e depois o conteúdo foi transferido para 5 ml de meio LIT suplementado com 10% (v/v) SFB e incubadas a 28 °C. Com 24 horas de transfecção, foi adicionado antibiótico de seleção G418 150 µg/ml. A morte total da cultura controle determina o fim do período de seleção dos parasitos. Paralelamente, os parasitos foram clonados em placa de 48 poços na presença de G418 150 µg/ml.

#### 4.2.12 Transfecção cassete de blasticidina em *T. cruzi* nas cepas CL-Brener e Y

A transfecção com o cassete de reparo por homologia contendo o gene de resistência da blasticidina ocorreu seguindo o mesmo protocolo realizado com o

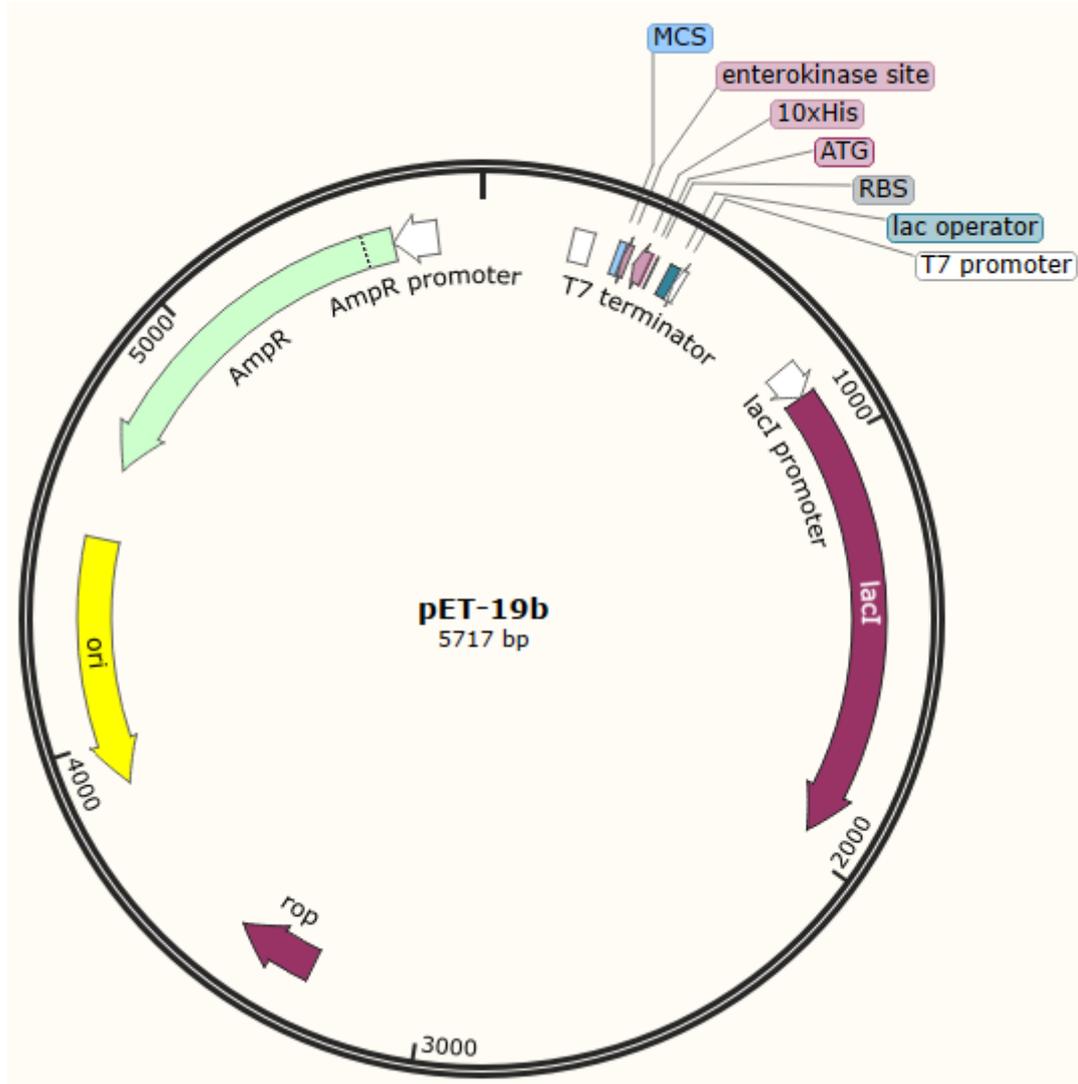
plasmídeo pTREX-n/Cas9/sgRNA. Somente a concentração de parasitos foi alterada para  $5 \times 10^6$  parasitos/ml. A transfecção se deu da seguinte forma: uma cultura transfectada com PBS para servir de controle, uma cultura transfectada somente com 25  $\mu\text{g}$  de pTREX-n/Cas9/sgRNA e uma cultura transfectada com a presença tanto do pTREX-n/Cas9/sgRNA e do cassete (1 $\mu\text{g}$ ). O controle foi separado em duas garrafas para servirem de controle para as 2 seleções. A seleção ocorreu com 48 horas, quando foi acrescido 150  $\mu\text{g}/\text{ml}$  de G418 às garrafas e a uma garrafa controle foi adicionado também 25  $\mu\text{g}/\text{ml}$  de blasticidina. Paralelamente, a clonagem dos parasitos foi realizada em placa de 48 poços com a adição dos antibióticos anteriormente citados.

#### 4.2.13 Expressão da catepsina B e análise por *Western Blot*

A região da pró-catepsina B foi sintetizada no vetor pET-19 B (**Figura 15**) para expressão da proteína feito pela empresa Genone. O plasmídeo na concentração de 50 ng foi transformado em *E. coli* nas cepas Rosetta (DE3) pLysS e BL21 (DE3). Após testes de melhores condições de expressão, os seguintes protocolos foram utilizados: três colônias foram aleatoriamente selecionadas e crescidas em 100 ml de meio LB contendo 100  $\mu\text{g}/\text{ml}$  de ampicilina à 37° C com rotação de 220 rpm até atingir a densidade óptica de 0,4 a 600 nm. A transformação na cepa Rosetta (DE3) pLysS seguiu as condições: IPTG 0,5mM (isopropil-1- $\beta$ -Dgalactopiranosídeo) por 5 horas com rotação de 220 rpm à 18°C. A transformação na cepa BL21 (DE3) foi induzida nas seguintes condições: IPTG 1mM por 3 horas com rotação de 220 rpm à 37° C. Após, a cultura foi centrifugada por 10 minutos à 10000 xg, o sobrenadante foi descartado e os pellets foram congelados à -20° C. No outro dia, os pellet foram lisados com 100  $\mu\text{l}$  BugBuster® (Novagen™) em cada tubo e levados para constante agitação durante 20 minutos. Posteriormente, as amostras foram centrifugadas por 10 minutos com rotação de 16000 xg à 4° C. O sobrenadante de cada amostra, que corresponde à fração solúvel, foi separado em outro tubo do pellet, fração insolúvel, e foram adicionados de tampão de amostra Laemmli (2% de SDS, 10% de glicerol, 5% de  $\beta$ -mercaptoetanol, 0,002% de azul de bromofenol, 125 mM de Tris HCl, pH 6.8) (LAEMMLI, 1970). Em seguida, foram fervidas por 5 minutos à 100° C e aplicadas em gel de poli-acrilamida SDS-PAGE com 12% (v/v) acrilamida/bis-acrilamida. A

transferência para membranas de nitrocelulose Amersham™ Protran™ (GE Healthcare™) ocorreu através do equipamento MiniProtean®3 Cell (BioRad™) corrente constante de 350 mA por 1 h. por 15 minutos. A membrana foi submersa em tampão de transferência Towbin (Tris 25 mM pH 8,3; glicina 192 mM; SDS 0,1% (w/v); e metanol 20% (v/v)) gelado, por 15 min antes do início da transferência.

Após, a membrana ser transferida, ela foi corada com solução Ponceau (Ponceau S 0,2% w/v; ácido acético 1%) por 5 minutos e lavada com água ultrapura. Então, foi bloqueada com solução contendo 5% (m/v) de leite desnatado em TBS (*Tris-Buffered Saline*, 50 mM Tris pH 7,5 e 150 mM NaCl) por 30 min, lavada duas vezes com TBS por 5 minutos e depois adicionado do anticorpo primário policlonal anti-His (Sigma Aldrich) diluído 1:2.000 em TBS em 0,1% Tween 20 (v/v) contendo 1% (m/v) de leite desnatado na diluição 1:2500 e incubada por 16 horas à 4° C. A membrana, então, foi lavada 5 vezes com TBS-T por 5 minutos cada lavagem e, assim, foi aplicado o anticorpo secundário (Goat Anti-Mouse IgG horseradish peroxidase conjugated – GAM) (Invitrogen™) diluído 1:30.000 em TBS-T e leite desnatado 1% (w/v) por 1 h à temperatura ambiente sob constante agitação. A membrana foi lavada 7 vezes por 5 minutos com TBS-T e a revelação ocorreu por quimioluminescência utilizando o substrato de peroxidase *ECL™ Prime Western Blotting Detection Reagent* (GE Healthcare™) no equipamento ImageQuant™ LAS 4000 (GE Healthcare™). As massas moleculares das proteínas foram estimadas por comparação com o padrão de massas moleculares *PageRuler™ Prestained Protein Ladder*, Invitrogen.



**Figura 15. Representação do plasmídeo pET-19b contendo suas regiões destacadas feito no SnapGene importado das sequências depositadas online do próprio programa. Na imagem está incluso o MCS, sítio de múltipla clonagem, onde a clonagem da pró-catepsina B foi realizada.**

#### 4.2.14 Purificação das frações solúvel e insolúvel da catepsina B

Após a indução, a fração solúvel expressa em BL21 (DE3) e a fração insolúvel expressa em Rosetta (DE3) pLysS foram purificadas seguindo as condições: as culturas foram centrifugadas por 20 minutos com rotação de 2000 G. O sobrenadante foi descartado e o pellet congelado. A lise do pellet foi feita com o uso de BugBuster® Protein Extraction Reagent (Sigma™) na proporção de 5 ml por cada grama de pellet e benzonase (25µg/ml), no caso da fração insolúvel foi adicionado o tampão de ligação (Tris 20 mM pH 8, NaCl 500 mM) na lise. O pellet foi ressuspensionado e mantido em agitação por 20 minutos. Então, as amostras foram centrifugadas a 4 °C por 10 min

com rotação de 16000 G. O sobrenadante corresponde à fração solúvel e o pellet à fração insolúvel. A purificação foi feita utilizando 500 µl da resina Ni Sepharose™ High Performance (GE Healthcare Life Sciences). A purificação da fração solúvel seguiu o seguinte protocolo: a amostra foi incubada 2 horas em uréia, seguida do equilíbrio da resina com tampão de ligação e as lavagens foram realizadas com o tampão e as concentrações crescentes 10, 20, 40, 60 e 80 mM de imidazol, respectivamente. Após, a eluição foi feita com 5 volumes de tampão de ligação com 250 mM de imidazol. A purificação da fração insolúvel seguiu o seguinte protocolo: o pellet foi ressuspendido com tampão de ligação e Ureia 8 M, assim como o equilíbrio da resina. A lavagem foi realizada com 30 volumes de tampão de ligação com Ureia 8 M e 10 mM de imidazol e a eluição foi feita com 7 volumes de tampão de ligação com Ureia 8 M e 250 mM de imidazol. As amostras então foram analisadas em gel SDS-PAGE 12% (v/v).

#### 4.2.15 Concentração das frações solúvel e insolúvel da catepsina B

As amostras foram concentradas utilizando o Amicon® Ultra-10 Centrifugal Filter Units (Merck™) sendo a coluna equilibrada com tampão Tris pH 7,0. A fração solúvel foi separada em dois concentrados, a primeira foi das amostras das lavagens, correspondentes às concentrações de imidazol até 80mM, e a segunda das amostras das eluições, correspondentes às cinco amostras com concentração de 250 mM de imidazol. Na fração insolúvel todas as eluições foram concentradas. As amostras foram centrifugadas por 10 minutos a 4 °C na rotação de 4000 G até concentrarem em 300 µl em ambas as concentrações. As amostras então foram analisadas em gel de poliacrilamida SDS-PAGE com 12% (v/v) acrilamida/bis-acrilamida.

# RESULTADOS

## 5 RESULTADOS

### PARTE I: DADOS GERADOS *IN SILICO*

#### 5.1 Localização dos alelos de catepsina B de *T. cruzi* na cepa CL Brener

De acordo com a pesquisa realizada na plataforma TriTrypDB (<https://tritrypdb.org/tritrypdb/app/>), os 2 alelos correspondentes ao gene *catbtc* na cepa CL Brener foram identificados. Como a cepa CL Brener tem origem híbrida, os alelos possuem origens filogenéticas diferentes e, portanto, o alelo proveniente da DTU TcI é identificado como *Esmeraldo-like* e pela letra S, possuindo o código TcCLB.510535.100. O outro alelo, proveniente da DTU TcII é identificado como *Non-Esmeraldo-like* e pela letra P com o código TcCLB.511827.100. Ambos os alelos estão localizados no cromossomo 13 (TcChr13) distribuídos em um homólogo constituinte do par, como também apresentando posições distintas em seus *loci* dos homólogos de origem (**Tabela 3**).

**Tabela 3-** Identificação e localização dos alelos do gene *catbtc*

ID da sequência do gene	Sequência genômica	Posição no cromossomo 13
TcCLB.510535.100	<i>catbtc_S</i>	473174 - 474175
TcCLB.511827.100	<i>catbtc_P</i>	473182 - 474183

Dados obtidos da plataforma TriTrypDB a partir do código de identificação de cada alelo <https://tritrypdb.org/tritrypdb/app/>.

#### 5.2 Alinhamento das sequências de nucleotídeos dos alelos da catepsina B de *T. cruzi* e desenho do sgRNA

O gene *catbtc* apresenta 1002 nucleotídeos para ambos os alelos conforme os dados obtidos no TriTrypDB (<https://tritrypdb.org/tritrypdb/app/>). Assim, o alinhamento entre estes alelos retornou que 979 nucleotídeos apresentam identidade entre si, o que representa 98% de identidade com 2% de polimorfismo. A região de 20 nucleotídeos destacada de vermelho entre as posições 331-350 foi observada como

uma região conservada entre os alelos e com uma localização propícia para a realização do corte pela *SpCas9*, portanto, foi escolhida como a sequência do sgRNA para ser posteriormente utilizada na estratégia de nocaute pelo sistema CRISPR-Cas9 (**Figura 16**).

```

catbtc_S      ATGCGTGTATTTTCTACTCAGCATCGCGCTTTTCTTTTCTGTTGTACGCCACGGCG
catbtc_P      ATGCGTGTATTTTCTACTCAGCATCGCTCTTTTCTTTTCTGTTGTACGCCACGGCG
*****

catbtc_S      GGACACTCTTTCCATGCGGAAGACGCTCCAATTTGACCGATGAGTTTTGGAACATGTG
catbtc_P      GGACACTCTTTCCATGCGGAAGACGCTCCAATTTGACAGATGAGTTTTGGAACATGTG
*****

catbtc_S      AACCGCTCAACGGCGGAAAGTGGACGGCTGGCCGCACAAGCCGAACAAAACACCTCAG
catbtc_P      AACCGTCTCAACGGCGGAAAGTGGACGGCTGGCCGCACAAGCCGAACAAAACACCTCAG
*****

catbtc_S      CGTCGCGGGGCAAGTCGGATGCTGGGGACTTTCCTACGAAACACCAGCATCTCCCCCG
catbtc_P      CGTCGCGGGGCAAGTCGGCTGCTGGGGACTTTCCTAAGAAACACCAGCATCTCCCCCG
*****

catbtc_S      CGGCAGTTCAGCGAGGAGGAGCTTCGTGTCCGCTCCAGGACAGGTTTGTATGCCGGAGAA
catbtc_P      CGGCAGTTCAGCGAGGAGGAGCTTCGTGAGCCGCTCCAGGACAGGTTTGTATGCCGGAGAA
*****

catbtc_S      GCGTGGCCGGAATGCCCCACCGTACCGGAATCCCGTGACCAATCGAGCTCGGCTCTTGC
catbtc_P      GCGTGGCCGGAATGCCCCACCATCACCGAATCCCGTGACCAATCGAGCTCGGCTCTTGC
*****

catbtc_S      TGGGCTGTGGCAGCAGCTTCAGCCATTTCCGACCGCTATTGCACCTTTGGTGGCGTACGG
catbtc_P      TGGGCTGTGGCAGCAGCTTCAGCCATTTCCGACCGCTATTGCACCTTTGGTGGCGTACGG
*****

catbtc_S      GACCTCAGAATTTCTGCGGGGATCTTATGTCGTGCTGCGATGTTTGGGTTTTGGTGC
catbtc_P      GACCTCAGAATTTCTGCGGGGATCTTATGTCGTGCTGCGATGTTTGGGTTATGGTGC
*****

catbtc_S      AATGGTGGCTATCCCGAAGTTGCCTGGGAGTACTACGCCGTGCACGGTATTGTTTCAGAG
catbtc_P      AATGGTGGCTATCCCGAAGTTGCCTGGGAGTACTACGCCGTGCACGGTATTGTTTCAGAG
*****

catbtc_S      TACTGCCAGCCTTATCCTTTTCCATCCTGTGCGCATCACGTGAACAGCAGCGATCTCTCC
catbtc_P      TACTGCCAGCCTTATCCTTTTCCATCCTGTGCGCATCACGTGAACAGCAGCGATCTCTCC
*****

catbtc_S      CCATGCTCCGGTGAATACGACACGCCTACGTGCAACAGCACTTGCACGGATAAGAAGATC
catbtc_P      CCATGCTCCGGTGAATACGACACGCCTACGTGCAACAGCACTTGCACGGATAAGAAGTGC
*****

catbtc_S      CCATTGATTAAGTACAGGGGAACACCTCATATGTCTTGAGCGGAGAGGACCTTTTAAG
catbtc_P      CCATTGATTAAGTACAGGGGAACACCTCATATCTCTTGAGCGGAGAGGAGTCTTTTAAG
*****

catbtc_S      CGCGAGTTAATTTGAATGGCCCATTCGAGGTTTCATTTTCAGTTTACGCTGACTTCTGTG
catbtc_P      CGAGAGTTACTTTGAATGGCCCATTCGAGGTTTCATTTTCAGTTTACGCTGACTTCTGTG
**

catbtc_S      GCCTACACAGGCGGCGTGTACAAGCATGTCGCTGGGATTTTCTCGGGGCCATGCGGTA
catbtc_P      GCCTACACAGGCGGCGTGTACAAGCATGTCGCTGGGACTTTTCTCGGGGCCATGCGGTA
**

catbtc_S      CGTATTGTTGGCTGGGTGAATTGAATGGAGAGCCGACTGGAAGATTGCGAACAGCTGG
catbtc_P      CGTATTGTTGGCTGGGTGAATTGAATGGAGAGCCGACTGGAAGATTGCGAACAGCTGG
*****

catbtc_S      AATCGTGAGTGGGGATGAATGGCTACTTCTTATTGCCCGTGGCGTGGATGAATGCGGA
catbtc_P      AATCGTGAGTGGGGATGAATGGCTACTTCTTATTGCCCGTGGCGTGGATGAATGCGGA
*****

catbtc_S      ATCGAGGAAGCGGCGTCCCGGTACGCCAAGAATACCATGA
catbtc_P      ATCGAGGAAGCGGCGTCCCGGTACTCCAAGAATACCATGA
*****

```

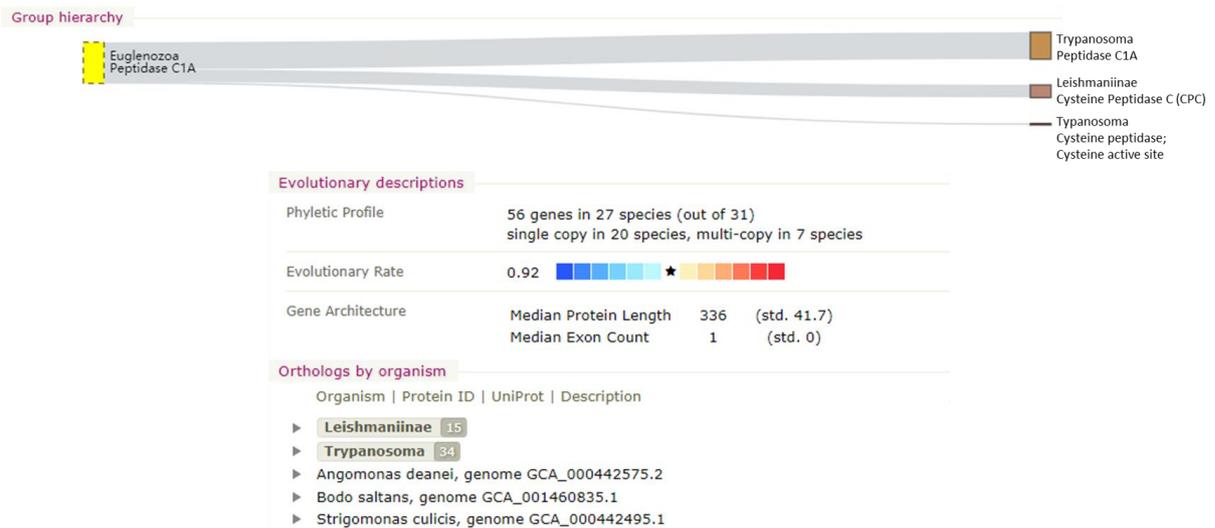
**Figura 16. Alinhamento dos alelos *Esmeraldo Like* e *Non-Esmeraldo Like*, respectivamente, da catepsina B de *Trypanosoma cruzi* na cepa CL Brener feito no MUSCLE. Os asteriscos indicam identidade entre os nucleotídeos de ambos os alelos e sua ausência, indica falta de identidade. A região vermelha destacada em vermelho indica a sequência de 20 nucleotídeos nas posições 331-350 escolhida para produzir o RNA guia.**

### 5.3 Alinhamento das sequências de aminoácidos dos alelos da catepsina B de *T. cruzi*

Ambos os alelos do gene *catbtc* apresentam 333 aminoácidos conforme os dados obtidos no TriTrypDB (<https://tritrypdb.org/tritrypdb/app/>). Assim, o alinhamento mostrou que 320 aminoácidos apresentam 100% de identidade entre si e que 328 aminoácidos possuem forte identidade com as propriedades bioquímicas apresentadas. As regiões principais entre os alelos também foram identificadas contendo um peptídeo sinal entre os resíduos 1-21, a região da pró-catepsina entre os resíduos 22-91 e a região correspondente à proteína madura e seu domínio Peptidase\_C1. Dentro desta região, é observada a presença do sítio catalítico constituído de cisteína, histidina e asparagina nas posições 120, 278, 298, respectivamente. Igualmente, é possível observar a presença do *loop* de oclusão, entre os resíduos 187-209. Nota-se que tanto a tríade catalítica quanto o *loop* de oclusão se apresentam conservados nos 2 alelos do gene *catbtc* (**Figura 17**).



apresentam uma taxa evolucionária de 92%. As proteínas codificadas têm em média 336 aminoácidos e 1 éxon. Os ortólogos são encontrados em 34 genes no gênero *Trypanosoma* com a presença do domínio Peptidase C1A; 15 genes da subfamília Leishmaniinae em que 9 deles estão no gênero *Leishmania* apresentando a Cisteína Peptidase C (CPC) como referência e os 7 restantes estão descritos nas espécies *Angomonas deanei*, *Bodo saltans* e *Strigomonas culicis* como cisteína peptidase e sítio ativo de cisteína (**Figura 18**).

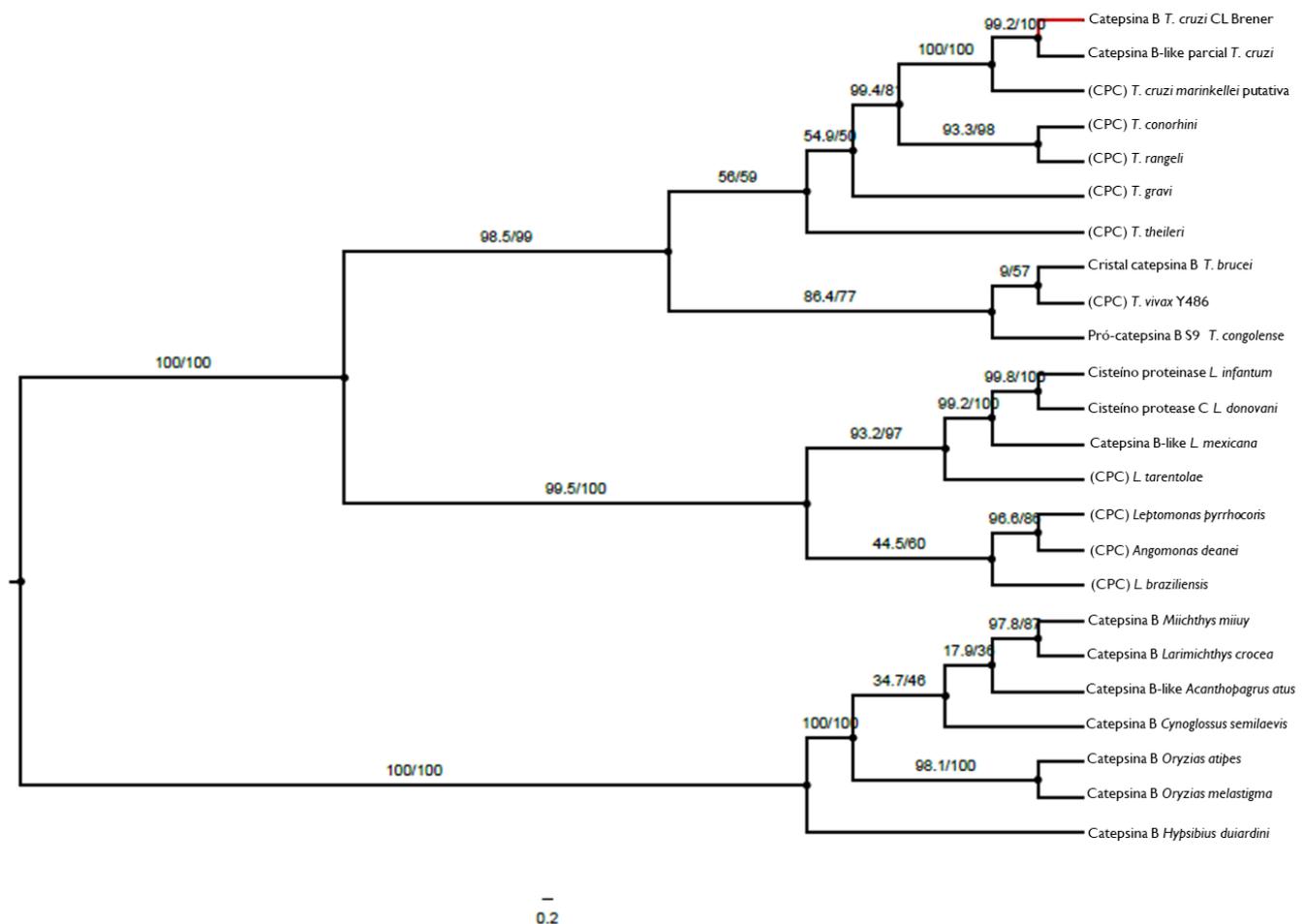


**Figura 18.** Descrição da ortologia do gene codificador da proteína catepsina B de *T. cruzi* da cepa CL Brener no filo Euglenozoa feito no OrthoDB.

#### 5.4.2 Árvore filogenética

Através de análise no BLAST-P com a sequência da catepsina B de *T. cruzi*, obteve-se 24 sequências de proteínas de diferentes organismos com alta porcentagem de identidade. Com esses dados, uma árvore filogenética foi gerada no servidor IQ-Tree Web Server (TRIFINOPOULOS et al., 2016) dispondo para cada ramo o valor de *alternate likelihood test (aLRT)* e *bootstrap*. Estes valores representam dados estatísticos aplicados à resolução da árvore filogenética. O *bootstrap* é um método para inferência estatística que se trata de uma medida de repetibilidade fornecendo um valor de confiança em relação à aparição de determinado ramo na árvore (TIBSHIRANI, 1993). O *aLRT* indica um ganho de probabilidade do ramo ao se manter invariável quando é aceita a hipótese nula e ele é considerado incorreto, enquanto o resto da árvore permanece sem modificações (ANISIMOVA; GASCUEL, 2006). A proteína XP\_812694.1, marcada de vermelho, representa a sequência

depositada da catepsina B de *T. cruzi*. Como esperado, a proteína AAL06325.1, *cathepsin B-like protease, partial* de *T. cruzi* aparece dividindo este mesmo ramo e a proteína EKF39488.1 *cysteine peptidase C (CPC), putative [Trypanosoma cruzi marinkellei]* na partição anterior. Os dados seguem aumentando a especiação entre as diversas espécies exibindo variadas espécies do gênero *Trypanosoma*, *Leishmania*, outros tripanossomatídeos, espécies modelo de peixes e o tardígrado, *Hypsibius dujardini* (Figura 19).



**Figura 19.** Árvore filogenética produzida a partir de 24 proteínas obtidas no BLAST com a catepsina B de *T. cruzi* gerada no IQ-Tree Web Server e visualizada no FigTree 1.4.4. São indicados os valores de *likelihood* e *bootstrap* gerados em cada nó. O nó referente à catepsina B de *T. cruzi* cepa CL Brener está destacado de vermelho. A linha na base da imagem indica o número total de mutações dividido pelo total de bases do alinhamento das sequências.

#### 5.4.3 Predição de massa molecular e ponto isoelétrico

Através do servidor ProtParam (HOCHSTRASSER, 2005), foi possível obter a predição da massa molecular de aproximadamente 37 kDa, ponto isoelétrico teórico de 5,74 com a presença de 333 aminoácidos na proteína completa (**Figura 20**).

```

      10      20      30      40      50      60
MRVYFSLSIA LFLFLLYATA GHSFHAEDAP ILTDEFLEHV NRLNGGKWA GRTSRTKHLT

      70      80      90     100     110     120
RRGASRMLGT FLRNTSILPP RQFSEEELRV PLQDRFDAGE AWPECPTVTE IRDQSSCGSC

     130     140     150     160     170     180
WAVAAASAI S DRYCTLGGVR DLRISAGDLM SCCDVCGFGC NGGYPEVAWE YYAVHGIVSE

     190     200     210     220     230     240
YCQYPFPFSC AHHVNSSDLS PCSGEYDTPT CNSTCTDKKI PLIKYRGNTS YVLSGEEPFK

     250     260     270     280     290     300
RELILNGPFE VSFSVYADFV AYTGGVYKHV AGIFLGGHAV RIVGWGELNG EPYWKIANSW

     310     320     330
NREWMNGYF LIARGVDECG IEGSGVAGTP RIP

```

[References and documentation](#) are available.

---

Number of amino acids: 333

Molecular weight: 36729.44

Theoretical pI: 5.74

**Figura 20.** Predição do número de aminoácidos, massa molecular e ponto isoelétrico teórico da *catbtc* completa codificada pelo alelo *Esmeraldo-like* feito no ProtParam.

#### 5.4.4 Presença de peptídeo sinal

A predição de sítio de clivagem de peptídeo sinal foi realizada pela plataforma SignalP 5.0 (ALMAGRO ARMENTEROS et al., 2019), mostrando que a proteína completa tem a possibilidade de 96,92% de possuir o sítio e que a probabilidade de ele ser clivado após o resíduo 21 é de 28,47%, indicando que o peptídeo sinal se encontra na porção inicial do gene (**Figura 21**).



**Figura 21.** Predição do sítio de clivagem de peptídeo sinal da *catbtc* completa codificada pelo alelo *Esmeraldo-like* feito no SignalP 5.0.

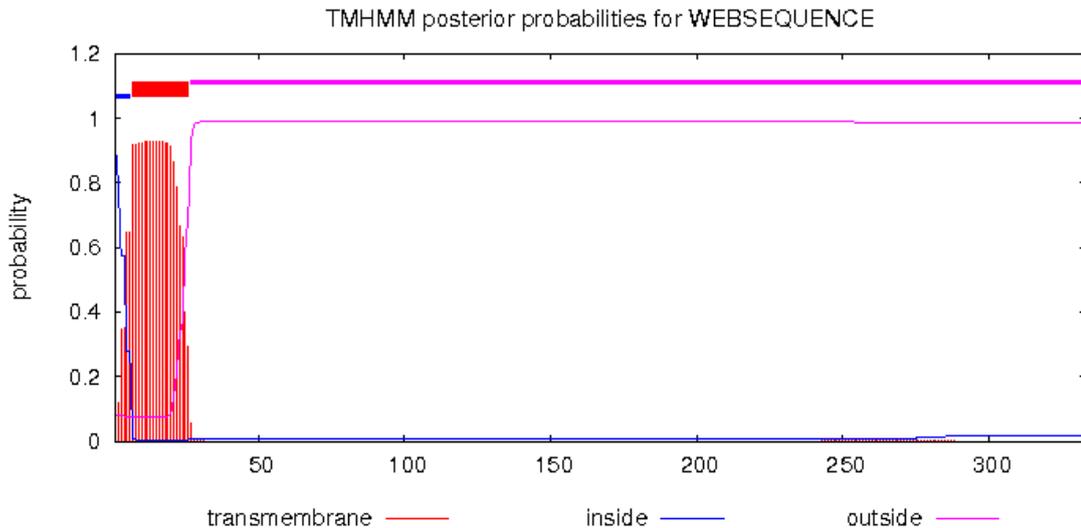
#### 5.4.5 Presença de domínio transmembrana

A predição de domínio transmembrana foi realizada pela plataforma TMHMM 2.0 (MÖLLER; CRONING; APWEILER, 2001), indicando a probabilidade de uma hélice transmembrana e que o número médio de aminoácidos dentro da hélice transmembrana é de 18,92. De acordo com o próprio servidor, se este valor for superior a 18, existe uma grande chance de haver um domínio transmembrana ou um sítio de clivagem de peptídeo sinal. O TMHMM 2.0 também previu o valor de 18,72 para o número médio de aminoácidos em hélices transmembranas nos primeiros 60 aminoácidos da proteína, valor do qual o programa sinaliza que a hélice transmembrana predita na região N-terminal da proteína pode significar um sítio de clivagem de peptídeo sinal. É calculado ainda, a probabilidade de que a região N-terminal da proteína está presente no lado citoplasmática da membrana, o qual representa 91,79% na catepsina B completa. Por último, é gerado um gráfico traçando a possibilidade de os resíduos estarem dentro, fora ou em hélice transmembrana, indicando os resíduos de 1-6 como dentro, 7-26 como hélice transmembrana e 27-333 como fora da membrana (**Figura 22**).

```

# WEBSEQUENCE Length: 333
# WEBSEQUENCE Number of predicted TMHs: 1
# WEBSEQUENCE Exp number of AAs in TMHs: 18.92119
# WEBSEQUENCE Exp number, first 60 AAs: 18.75446
# WEBSEQUENCE Total prob of N-in: 0.91795
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 inside 1 6
WEBSEQUENCE TMHMM2.0 TMhelix 7 26
WEBSEQUENCE TMHMM2.0 outside 27 333

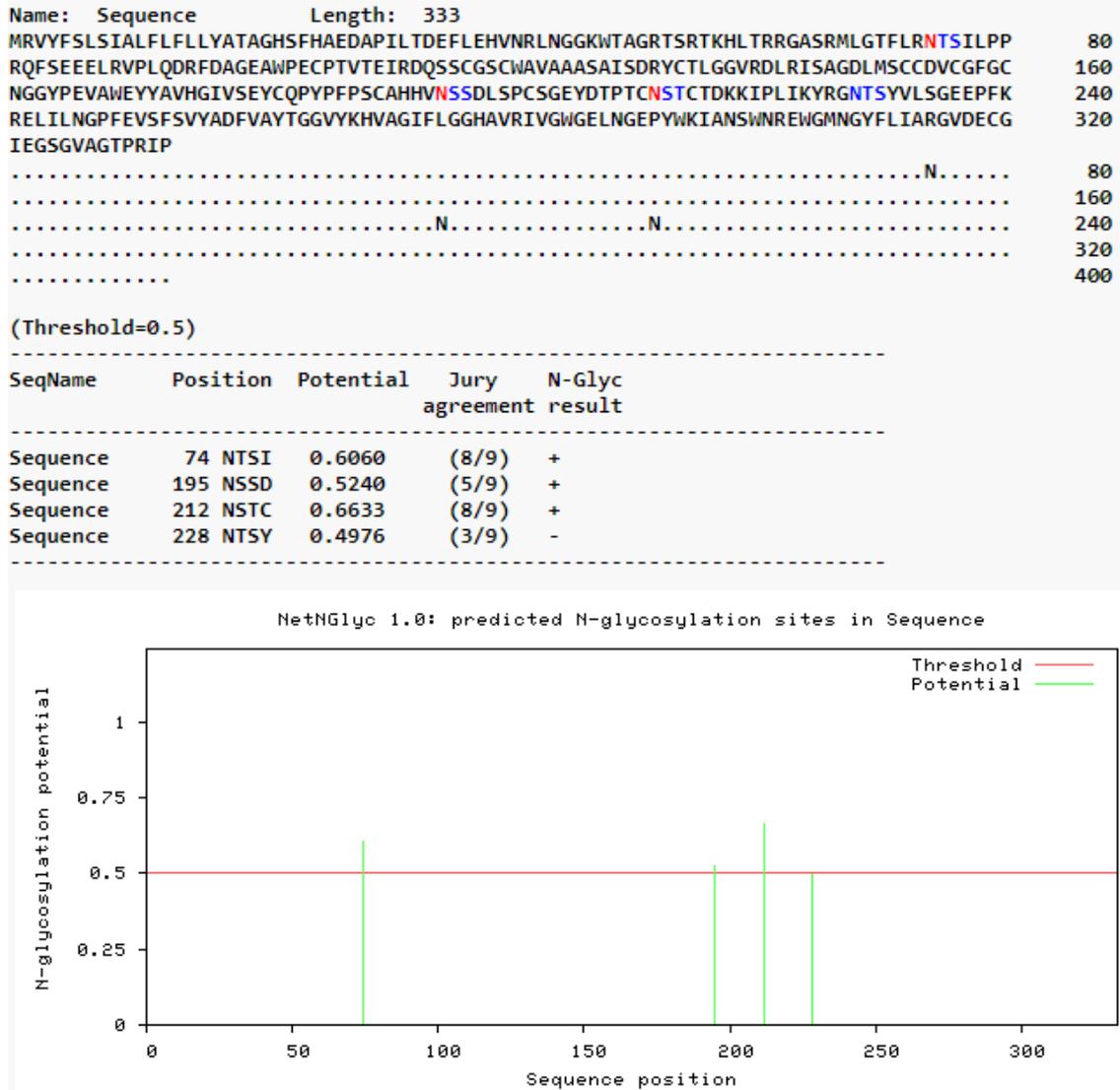
```



**Figura 22. Predição do domínio transmembrana da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no TMHMM 2.0.**

#### 5.4.6 Motivos de N-glicosilação

A predição de motivos de N-glicosilação foi realizada pela plataforma NetNGlyc 1.0 (GUPTA; BRUNAK, 2002). O programa indica os potenciais resíduos de asparagina (N) glicosilados, mostrando a sequências de asparagina - aminoácido X - serina e/ou treonina (N-Xaa-S/T) observadas. Essa sequência é característica do processo de N-glicosilação, exceto quando o aminoácido Xaa é prolina, pois este resíduo torna a asparagina inacessível ao processo. Porém, a presença da sequência nem sempre é indicação de um sítio. Como resultado, 4 potenciais sequências foram verificadas nas posições 74, 195, 212 e 228, mas apenas as 3 primeiras foram preditas como motivos de N-glicosilação (asparaginas vermelhas). O limiar de 0,5 é configuração do programa que o gera através de uma média de resultados provenientes de 9 servidores neurais, dado por meio do *jury agreement* (**Figura 23**).



**Figura 23. Predição de motivos de N-glicosilação da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no NetNGlyc 1.0. Motivos potenciais (azul) e aminoácidos preditos (vermelho) como sítios de N-glicosilação,**

#### 5.4.7 Motivos de O-glicosilação

A predição de motivos de O-glicosilação foi realizada pela plataforma NetOGlyc 4.0 (STEENTOFT et al., 2013), sendo capaz de prever 14 possíveis motivos de O-glicosilação na proteína completa nas posições 53, 56, 60, 65, 75, 76, 84, 189, 196, 197, 200, 203, 208 e 210, os quais obtiveram uma pontuação maior que 0,5 de acordo com os parâmetros do servidor (**Figura 24**).

```

##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 21-7-9
##Type Protein
#seqname      source  feature start  end  score  strand  frame  comment
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 6    6    0.00530279 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 8    8    0.00769328 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 19   19   0.0332187 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 23   23   0.121482 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 33   33   0.0203492 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 49   49   0.458873 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 53   53   0.511256 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 54   54   0.355972 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 56   56   0.563313 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 60   60   0.707815 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 65   65   0.783707 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 70   70   0.335052 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 75   75   0.777092 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 76   76   0.902849 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 84   84   0.538231 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 107  107  0.388611 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 109  109  0.14949 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 115  115  0.0277498 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 116  116  0.0358952 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 119  119  0.049224 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 127  127  0.0580211 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 130  130  0.0623665 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 135  135  0.109094 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 145  145  0.371029 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 151  151  0.0223335 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 179  179  0.0839565 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 189  189  0.53014 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 196  196  0.646544 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 197  197  0.739665 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 200  200  0.714907 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 203  203  0.767152 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 208  208  0.5 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 210  210  0.561202 .    . #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 213  213  0.249333 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 214  214  0.308012 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 216  216  0.290692 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 229  229  0.0615239 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 230  230  0.0383247 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 234  234  0.00542474 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 252  252  0.00841882 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 254  254  0.00576156 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 263  263  0.0443075 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 299  299  0.00587375 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 324  324  0.0287704 .    .
SEQUENCE     netOGlyc-4.0.0.13 CARBOHYD 329  329  0.0743503 .    .

```

**Figura 24. Predição de motivos de O-glicosilação da catepsina B completa codificada pelo alelo *Esmeraldo-like* feito no NetOGlyc 4.0.**

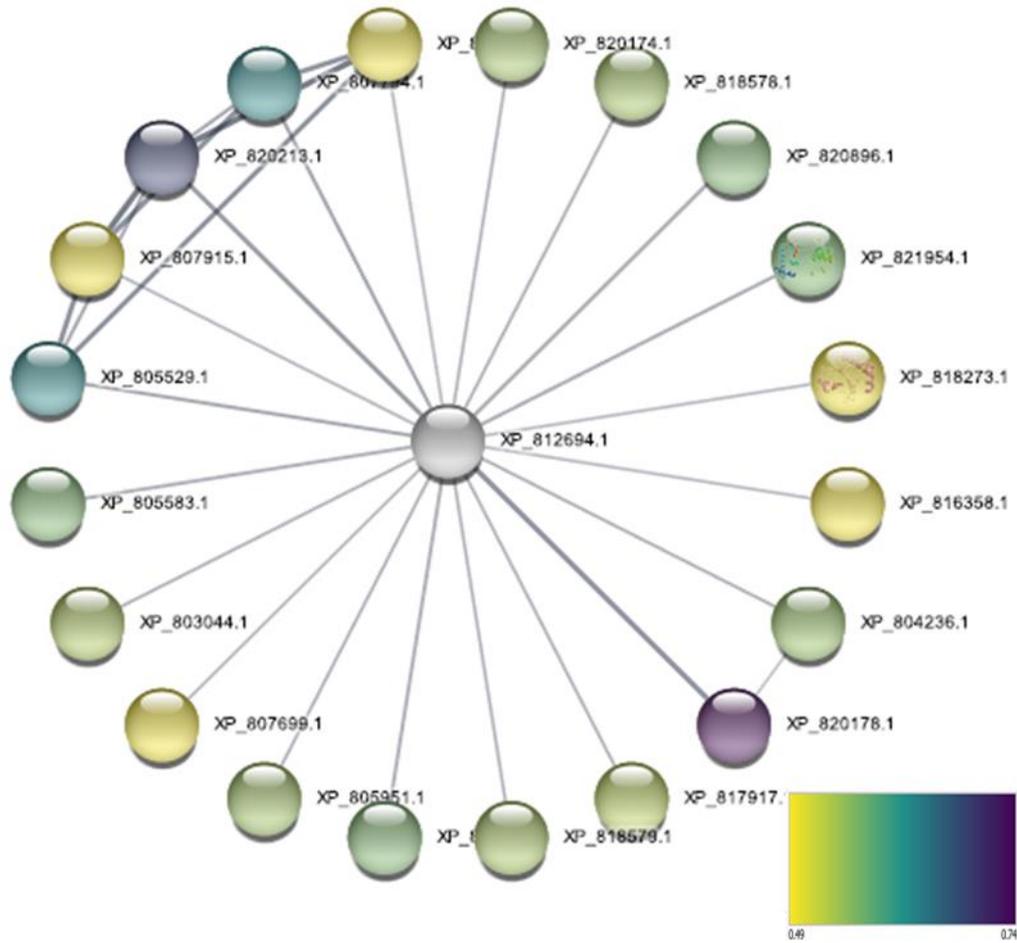
#### 5.4.8 STRING

A predição de interação proteína-proteína foi realizada pelo servidor STRING (SZKLARCZYK et al., 2019) entre a catepsina B completa codificada pelo alelo *Esmeraldo-like* e as descrições na literatura com as proteínas relacionadas ao *T. cruzi*. Os parâmetros utilizados foram *full network* que mostra tanto associações físicas como funcionais de proteínas, interações de confiança com uma pontuação média de confiança na interação (0,4) e com no máximo 20 interações. Então, 21 nós retornaram em um total de 29 pontas de interações (**Figura 25**). As interações relacionadas à catepsina B humana (CTSB) também foram avaliadas. Os parâmetros usados foram *full network* com interações de confiança dispendo a pontuação mais

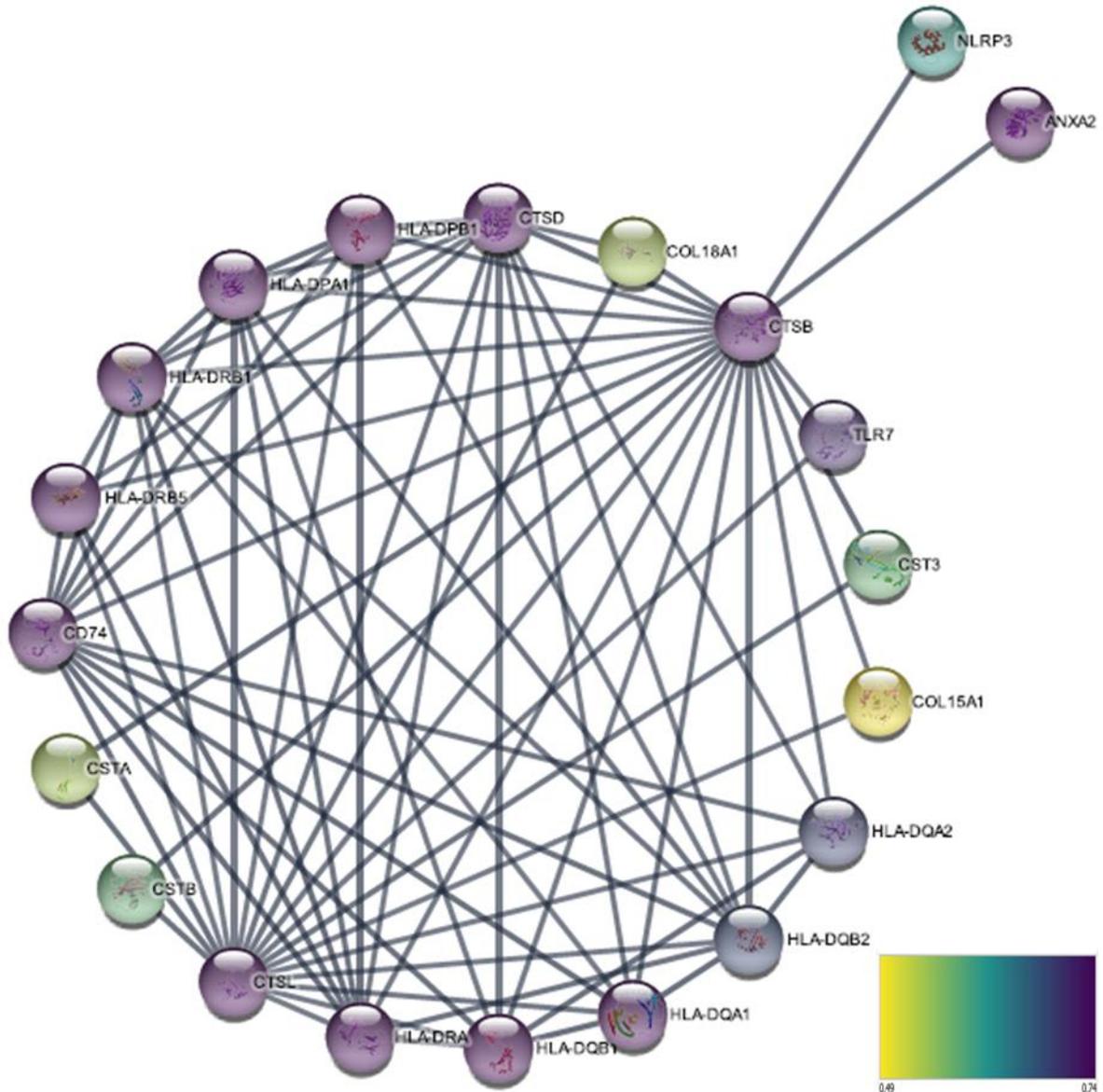
alta de confiança na interação (0,9) apresentando no máximo 20 interações. O resultado obtido foi de 21 nós/nichos/*clusters* com um total de 76 pontas de interação (**Figura 26**). Então, ambos resultados foram submetidos no programa Cytoscape (IDEKER, 2003), que tem capacidade de integrar com o banco de dados do STRING realizando o enriquecimento dos dados. A organização do STRING do *T. cruzi* foi feita pelo *score* de interação com as outras proteínas variando em um mapa contínuo de amarelo a roxo, de 0,49 a 0,74, respectivamente. O STRING de *H. sapiens* foi organizado de acordo com a presença no lisossomo, variando também em um mapa contínuo de amarelo a roxo, com o *score* de 1,06 a 5,0, respectivamente. Ambas as interações foram adaptadas para o layout circular.

Das interações com maior *score* e confiança da catepsina B (XP\_812694.1), localizada no centro da **Figura 25**, com o *T. cruzi*, é possível observar a cruzipaína (XP\_820178.1) e a subunidade  $\beta$ -like da proteína ligante do nucleotídeo guanina (XP\_820213.1). Ainda mostra relação alta com as syntaxinas (XP\_805529.1 e XP\_807794.1) que, por sua vez, apresentam interação de confiança com proteínas de organização e compactação do Golgi, homólogas da GRASP humana (XP\_807915.1 e XP\_820925.1). Sobre as outras interações que a proteína realiza, 7 delas correspondem a descrições de cisteíno peptidases (XP\_820174.1; XP\_818578.1; XP\_821954.1; XP\_805584.1; XP\_805951.1; XP\_803044.1; XP\_805583.1), 3 relacionadas a proteínas putativas da cruzipaína (XP\_820896.1; XP\_804236.1; XP\_818579.1), 2 que contêm domínio de tioredoxina (XP\_818273.1; XP\_816358.1), 1 proteína dissulfeto isomerase putativa (XP\_807699.1) e 1 Pitrilisina-like metalopeptidase (XP\_817917.1).

Em relação à catepsina B de *H. sapiens* (CTSB) é possível verificar que a maioria das proteínas que ela possui interação estão altamente relacionadas ao lisossomo com exceção das cistatinas A, B e C (CSTA, CSTB, CST3, respectivamente), as cadeias- $\alpha$ -1 dos tipos XV e XVIII (COL15A1 e COL18A1, respectivamente). Das interações com proteínas com alto *score* de presença nos lisossomos, 12 se referem à resposta imune (MHCs de classe 2 <HLA>, CD74, TLR7 e NLRP3), além da presença da pró-catepsina L (CTSL) e catepsina D (CTSD) e anexina 2 (ANXA2). Nas interações de alta confiança, destacam-se o MHC de classe 2 DQ $\beta$ 2 (HLA-DQB2), anexina A2 (ANXA2), o inflamassoma NLRP3 e a cistatinas A, B e C (CSTA, CSTB, CST3, respectivamente), na **Figura 26**.



**Figura 25. Interações realizadas pela catepsina B descritas com proteínas relacionadas ao *T. cruzi* feito no STRING. As cores indicam o score de interação com as outras proteínas variando em um mapa contínuo de amarelo a roxo. As linhas mais grossas apontam interações com maior grau de confiança.**

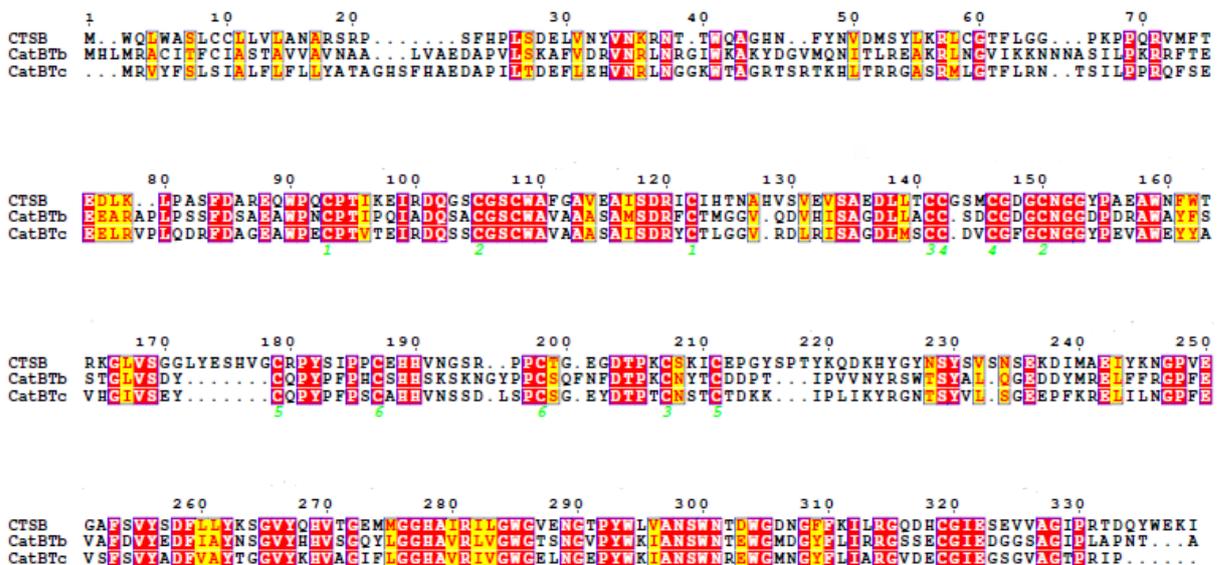


**Figura 26. Interações realizadas pela catepsina B descritas com proteínas relacionadas a *Homo sapiens* feito no STRING.** As cores indicam a presença de interação com proteínas descritas no lisossomo, variando em um mapa contínuo de amarelo a roxo. As linhas mais grossas apontam interações com maior grau de confiança.

#### 5.4.9 Alinhamento de sequências de catepsina B de diferentes organismos

O alinhamento da catepsina B de *T. cruzi* (CatBTc) e as sequências de catepsina B de *T. brucei* (CatBTb) e *H. sapiens* (CTSB) foi realizado pelo servidor ESPript 3.0 (ROBERT; GOUET, 2014), o qual compara dados de sequência similares e de estrutura secundária através do alinhamento com algum PDB fornecido. Então, as três sequências foram alinhadas e comparadas com o PDB 1GMV. Nota-se a presença de uma lacuna nas sequências de CatBTb e CatBTc em relação a CTSB

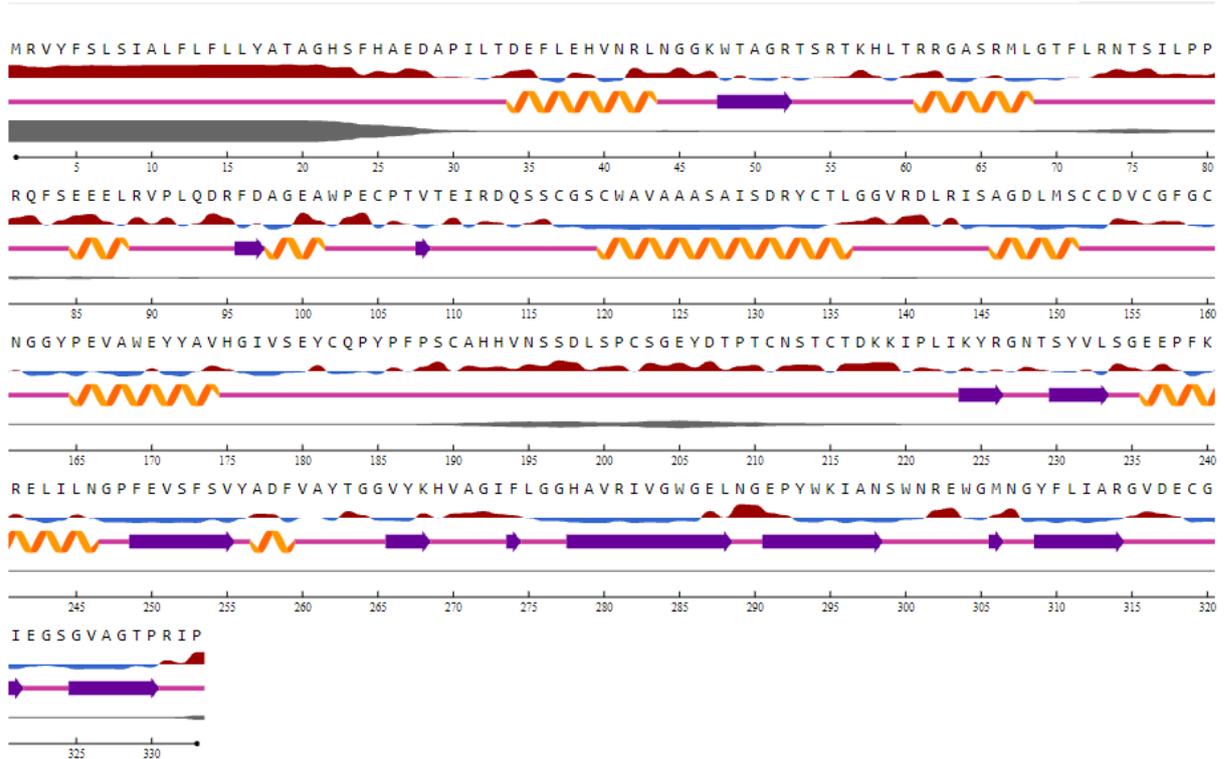
nas posições 172-179 (**Figura 27**). Verifica-se que as três sequências possuem um alto grau de identidade, principalmente na porção madura que começa a partir do resíduo 77. O sítio catalítico (resíduos 108, 278 e 298), os resíduos de histidina importantes para a manutenção do *loop* de oclusão (resíduos 189 e 190) e as seis cisteínas formadoras de pontes de dissulfeto se encontram conservadas nas três sequências.



**Figura 27. Alinhamento das sequências de catepsina B de *H. sapiens* (CTSB), *T. brucei* (CatBTb) e *T. cruzi* (CatBTc).** As cores vermelhas indicam resíduos que possuem 100% de identidade nas três sequências, as cores amarelas, fortes propriedades bioquímicas similares e os números de verde, as pontes de dissulfeto com seus respectivos pares. O sítio ativo está presente nos resíduos 108, 278 e 298.

#### 5.4.10 Predição de estrutura secundária da catepsina B de *T. cruzi*

O servidor ESPrift 3.0 tem capacidade de prever a estrutura secundária por um PDB fornecido. Logo, a estrutura referente à catepsina B de *T. cruzi* não pode ser criada pelo programa. Então, o servidor NetSurfP-2.0 (KLAUSEN et al., 2018) foi utilizado para realizar a predição a partir da estrutura primária da catepsina B de *T. cruzi* (**Figura 28**). Em resumo, a predição indica 9  $\alpha$ -hélices e 14 folhas  $\beta$ .



**Figura 28.** Predição da estrutura secundária da catepsina B de *T. cruzi* a partir da sequência primária no servidor NetSurfP-2.0. Na imagem, as voltas laranjas indicam regiões de predição de  $\alpha$ -hélice; as setas roxas, folhas  $\beta$ ; as linhas vermelhas, exposição à superfície; as linhas azuis, não-exposição à superfície e linhas cinzas, o nível de desordem dos resíduos pela espessura da linha.

#### 5.4.11 Modelagem da estrutura tridimensional da catepsina B de *Trypanosoma cruzi*

Os seis modelos 3D da catepsina B de *T. cruzi* foram gerados utilizando quatro servidores diferentes. Os modelos foram refinados, utilizando o servidor GalaxyREFINE (KO et al., 2012), gerando um total de 12 modelos. Através do servidor Swiss-Model (SCHWEDE et al., 2003) foram obtidos dois modelos utilizando PDBs moldes diferentes: o primeiro do PDB 4HWY (CHO, 2012) (estrutura apo da procathepsin B de *Trypanosoma brucei*) com identidade sequencial de 60%, cobertura de 91,9% em comparação à sequência depositada e um QMEAN de 0,83. O segundo modelo utilizou como molde o PDB 3HHI (KERR et al., 2010) (estrutura holo da cathepsin B de *T. brucei* complexada ao inibidor CA074) que obteve 60,39% de identidade sequencial, 77% de cobertura e um QMEAN de 0,85. O modelo obtido com o PDB 4HWY foi submetido no servidor Molprobitry com *Ramachandran favored* de 95,41%, *Clashscore* de 0 e *Molprobitry score* de 0,97, com percentil de 100% para ambos parâmetros. O modelo obtido com o PDB 3HHI apresentou *Ramachandran*

*favored* de 94,49%, *Clashscore* de 0,79 e *Molprobability score* de 1,12 com percentil de 99% para ambos os parâmetros.

Através do servidor I-Tasser, foram gerados três modelos: (i) modelo sem restrições ou molde específico; (ii) modelo utilizando como molde o PDB. 4HWY e (iii) modelo utilizando como molde o PDB 3HHI. Todos os modelos precisaram de 2 refinamentos para alcançar scores bons dentro dos parâmetros. O modelo gerado pela configuração padrão (sem restrições ou molde específico) apresentou *Ramachandran favored* de 96,45%, *Clashscore* de 8,5 e *Molprobability score* de 1,69, com percentil de 79% e 89%, respectivamente. O modelo que teve como molde o PDB 4HWY obteve *Ramachandran favored* de 96,77%, *Clashscore* de 5,53 e *Molprobability score* de 1,5, com o percentil de 92% e 95%, respectivamente. Enquanto o modelo com base no PDB 3HHI atingiu o *Ramachandran favoured* de 96,13%, *Clashscore* de 7,01 e *Molprobability score* de 1,65, com o percentil de 87% e 91%, respectivamente.

Posteriormente, o trRosetta foi utilizado para a modelagem de estrutura tridimensional da proteína. A sequência da pró-catepsina foi utilizada, em virtude da dificuldade da proteína se enovelar corretamente sem esta região. O modelo trRosetta apresentou TM-score de 0,89, indicando alta confiabilidade, além de *Ramachandran favored* de 96,13%, *Clashscore* de 2,98 e *Molprobability score* de 1,35, ambos com 98% do percentil avaliado.

Por último, o AlphaFold, que dispõe de um conjunto de modelos 3D de proteínas prontos, foi utilizado para obter um modelo da proteína a partir do código do Uniprot Q4DE83 correspondente à catepsina B de *T. cruzi* da cepa CL Brener. Por se tratar da sequência primária completa, o modelo da proteína possui todas as porções: pré, pró e madura da protease. Em relação à qualidade do modelo, entre os resíduos 1-27 iniciais, os aminoácidos receberam a pontuação de qualidade entre muito baixo e baixo. Os demais resíduos obtiveram resultado de confiável ou muito alto, onde a maioria se mostrou muito alto, principalmente na região madura da proteína. Em relação aos parâmetros observados pelo *Molprobability*, *Ramachandran favored* de 96,68%, *Clashscore* de 0,4 e *Molprobability score* de 1,16, ambos com percentil de 99%. Os parâmetros de qualidade dos modelos trRosetta e AlphaFold refinados, se mostraram piores em relação aos originais. Evidenciando que o refinamento estrutural não é interessante para estruturas tridimensionais preditas utilizando métodos de inteligência artificial. Na **Tabela 4**, comparou-se parâmetros de qualidade dos modelos originais e refinados obtidos com os quatro servidores utilizados. As cores para cada

parâmetro destacam se estão dentro do limite aceitável (verde), levemente acima do limite (amarelo) e fora do limite aceitável (vermelho). O modelo que teve como molde o PDB 4HWY feito pelo servidor Swiss-Model foi escolhido entre os gerados. Porém, apesar de dispor dos melhores parâmetros entre os outros modelos obtidos, esse modelo não possuía todos os aminoácidos correspondentes à sequência primária da proteína. Dessa forma, uma vez que os modelos originais obtidos com os servidores trRosetta e AlphaFold utilizam o mesmo método de modelagem e alcançaram bons parâmetros de qualidade e semelhanças, ambos foram selecionados para seguir as próximas etapas do estudo.

**Tabela 4 - Modelos 3D da CatB de *T. cruzi***

Modelos	SM Original	SM Refinado	I.T. Padrão Original	I.T. Padrão Refinado	I.T. 4HWY Original	I.T. 4HWY Refinado	I.T. 3HHI Original	I.T. 3HHI Refinado	tr Original	tr Refinado	AF Original	AF Refinado
<i>Molprobit</i> Score	0,97	1,28	2,78	1,69	2,41	1,50	2,68	1,65	1,35	1,48	1,16	1,34
<i>Clashscore</i>	0	5,19	5,54	8,5	3,62	5,53	5,53	7,01	2,98	8,92	0,4	6,14
<i>Ramachandran</i> <i>favored</i>	95,41%	98,03%	82,26%	96,45%	84,19%	96,77%	83,55%	96,13%	96,13%	99,03%	96,68%	98,19%

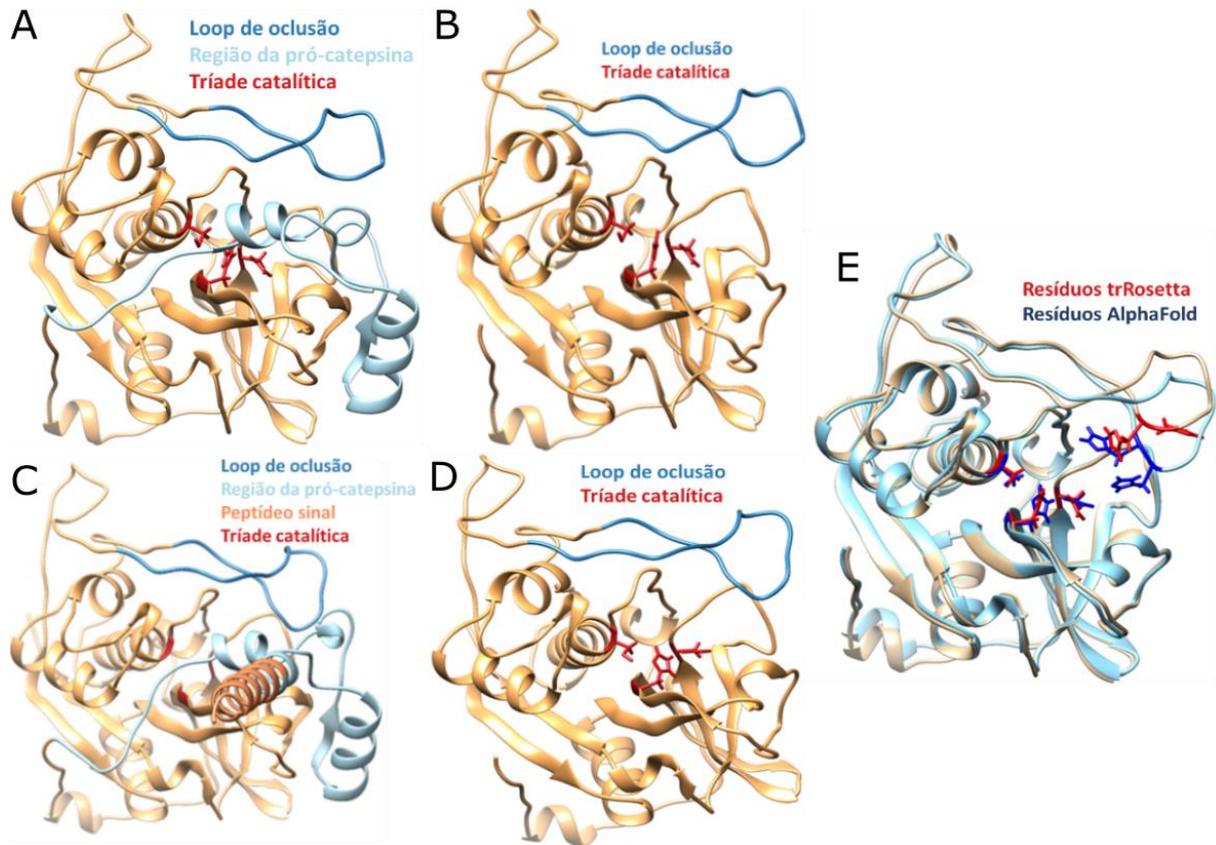
**Modelos referentes ao Swiss-Model (SM), I-Tasser (I.T.), trRosetta (tr) e AlphaFold (AF).** No modelo I-Tasser foram obtidas três versões: configuração padrão, a partir do molde PDB 4HWY e do PDB 3HHI. Os valores e cores dos parâmetros de qualidade foram extraídos do servidor *Molprobit*, sendo com verde para valores dentro do limite aceitável; amarelo, levemente acima do limite; e vermelho, fora do limite aceitável.

Por motivos de análise e comparação, as regiões pré e pró foram cortadas de ambos os modelos. Além disso, a presença da região pró na estrutura impediria o *docking* de pequenas moléculas no sítio catalítico. Então, o modelo gerado pelo trRosetta (**Figura 29 A**) da catepsina B de *T. cruzi* se constituiu de 312 aminoácidos, dispondo de uma região da pró-catepsina entre os resíduos 1-56, para então adentrar a região correspondente a proteína madura e seu domínio peptidase\_C1 percorrendo os resíduos 57-312. Dentro da região madura, a tríade catalítica se encontra nas posições: Cys99, His257 e Asn277, o *loop* de oclusão, na região entre os resíduos Phe166-Pro188.

A partir desse modelo, retirou-se a região da pró-catepsina (**Figura 29 B**), resultando em uma estrutura contendo apenas a porção madura da protease. O

modelo gerado pelo AlphaFold (**Figura 29 C**) diferentemente do modelo anterior, possui o peptídeo sinal. Portanto, o modelo construído possui 333 aminoácidos, com a região pré compondo os primeiros 21 resíduos, seguida da pró-catepsina entre os aminoácidos 22-77 e com os resíduos restantes compreendendo a porção madura. O sítio catalítico se encontra nas posições Cys120, His278 e Asn298. O *loop* de oclusão está entre os resíduos Phe187-Pro209. Da mesma forma do trRosetta, foi gerado um modelo retirando as regiões pré e pró com 256 aminoácidos (**Figura 29 D**).

É possível observar a presença do *loop* de oclusão marcado de azul, a tríade catalítica de vermelho, a região da pró catespina de azul claro e o peptídeo sinal de marrom com a representação dos modelos gerados tridimensionalmente originalmente (**Figuras 29 A e 29 C**). A sobreposição dos modelos também foi feita (**Figura 29 E**). Observou-se poucas diferenças entre os modelos e o RMSD entre ambas as estruturas foi de 2,3 Å (calculado considerando todos os átomos exceto hidrogênios). A maior diferença está relacionada a conformação do *loop* de oclusão. No modelo trRosetta, o *loop* adota uma conformação mais aberta com o resíduo His192 voltado para o solvente e exposição maior do sítio catalítico, em comparação ao modelo do AlphaFold (**Figura 29 E**).



**Figura 29. Modelos tridimensionais da catepsina B de *Trypanosoma cruzi*.** **A.** Pró-catepsina B, gerada pelo servidor trRosetta. **B.** Catepsina B madura, gerada pelo servidor trRosetta. **C.** Catepsina B completa, gerada pelo servidor AlphaFold. **D.** Catepsina B madura, gerada pelo servidor AlphaFold. **E.** Sobreposição das catepsina B maduras, geradas pelos servidores trRosetta (bege) e AlphaFold (azul). Os resíduos do sítio catalítico e *loop* de oclusão (representados em *sticks*) dos modelos trRosetta e AlphaFold foram destacados em vermelho e azul, respectivamente. No centro da imagem se encontram os resíduos do sítio catalítico (Cys120, His278 e Asn298) e na parte superior, os do *loop* de oclusão, com destaque para os resíduos de histidina (His192 e His193) importantes para manutenção da interação com ligantes. Figuras geradas no programa UCSF CHIMERA.

#### 5.4.12 Validação do protocolo de *docking* com compostos ativos e *decoys*

A partir dos 14 compostos ativos de CatB de *T. brucei*, *S. mansoni* e *H. sapiens*, encontrados na literatura (reportados previamente na **Fig. 11**), geramos 504 *decoys* e realizamos os *dockings* nos programas DockThor e Glide. Os 518 compostos (14 ativos e 504 *decoys*) resultantes da preparação dos ligantes nos modelos trRosetta e AlphaFold foram validados pelo parâmetros obtidos após o *docking* no DockThor e Glide. A partir desses resultados, foram avaliados métricas em 3 diferentes populações desses dados: *top* 1%, *top* 5% e *top* 10%, representando os primeiros 1%,

5% e 10% compostos dos resultados, respectivamente, de acordo com o *score* obtido para calcular a eficiência deste método.

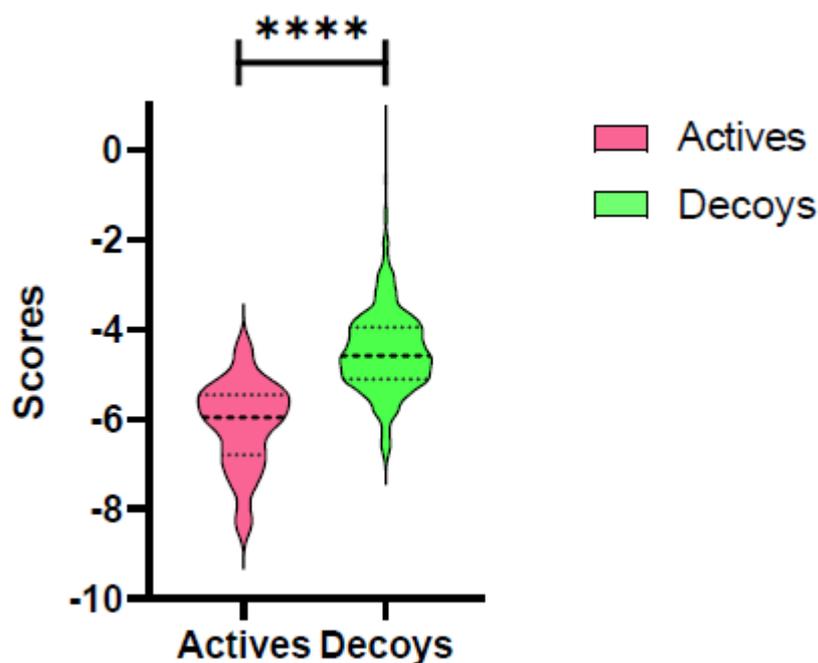
Durante a validação, o modelo trRosetta não obteve boas métricas em nenhum pH testado. Portanto, apenas o modelo AlphaFold foi considerado. Deste modelo, o *docking* pelo programa Glide no pH 4,5 *top* 10% se mostrou com as melhores métricas e, por consequência, foi selecionado para dar continuidade nos experimentos com os compostos do MNHN.

**Tabela 5** - Métricas de validação dos cálculos de *docking* dos compostos ativos e *decoys* nos modelos CatBTC (trRosetta e AlphaFold), realizados nos servidores DockThor e Glide.

	Top 1%						Top 5%						Top 10%					
	AUC	CCR	Se	Sp	EF	BEDROC	CCR	Se	Sp	EF	BEDROC	CCR	Se	Sp	EF	BEDROC		
DockThor trRosetta	0.76	0.6	0.59	0.61	20.0	0.63	0.62	0.58	0.65	6.77	0.41	0.64	0.61	0.66	4.0	0.41		
DockThor AlphaFold	0.64	0.54	0.54	0.53	8.87	0.23	0.56	0.53	0.6	5.33	0.27	0.56	0.53	0.6	2.67	0.29		
Glide trRosetta	0.69	0.53	0.5	0.56	13.33	0.44	0.56	0.5	0.62	6.67	0.33	0.56	0.5	0.63	3.33	0.35		
Glide AlphaFold	0.91	0.68	0.73	0.63	22.4	0.68	0.7	0.72	0.68	9.33	0.5	0.75	0.77	0.73	6.0	0.56		
DockThor trRosetta pH 7,4	0.53	0.5	0.5	0.5	6.67	0.11	0.5	0.5	0.5	1.33	0.1	0.5	0.5	0.51	1.33	0.13		
Glide trRosetta pH 7,4	0.82	0.57	0.5	0.65	26.67	0.74	0.6	0.5	0.69	9.33	0.5	0.64	0.56	0.73	6.0	0.55		
Valores referência	[>0.7]	[0-1]	[0-1]	[0-1]	[0-100]	[0-1]	[0-1]	[0-1]	[0-1]	[0-20]	[0-1]	[0-1]	[0-1]	[0-1]	[0-10]	[0-1]		

Na **Tabela 5**, a comparação das métricas nas populações testadas computadas pelo programa Glide no modelo AlphaFold pH 4,5, que assinala a população *top* 10% como a melhor amostra. Obteve-se quatro das cinco métricas analisadas consideradas boas. Na **Figura 30**, o gráfico violino foi gerado em função do *score* de *docking* obtido pelo Glide sobre o modelo gerado no AlphaFold pH 4,5, onde se observa que houve diferença significativa, com valor de  $P = < 0,0001$ , comparando os *scores* preditos para os ativos (rosa) e *decoys* (verde) após avaliação por teste Mann-Whitney. A linha preta central mais grossa representa a mediana de cada grupo e as linhas pretas mais finas superiores e inferiores representam o 3º quartil e 1º quartil de cada grupo, respectivamente. Os quartis separam os valores do conjunto em quatro partes iguais, logo, os 1º e 3º quartis delimitam os 25% dos menores valores e os 25%

dos maiores valores, respectivamente. Nota-se que os ativos obtiveram scores mais negativos (energia de ligação mais favorável) em relação aos *decoys*.



**Figura 30. Gráfico violino da distribuição dos scores de *docking* dos compostos ativos (rosa) e *decoys* (verde), com valor exato e significativo de  $P = < 0,0001$ . Foi aplicado o teste não paramétrico *Mann-Whitney*, apresentando mediana dos ativos em -5,962 e  $n = 15$ ; enquanto a mediana dos inativos foi de -4,584 e  $n = 521$ . A linha preta central mais grossa representa a mediana e as linhas pretas mais finas superiores e inferiores representam o 3º quartil e 1º quartil, respectivamente.**

#### 5.4.13 Triagem virtual dos compostos do Museu Nacional de História Natural de Paris

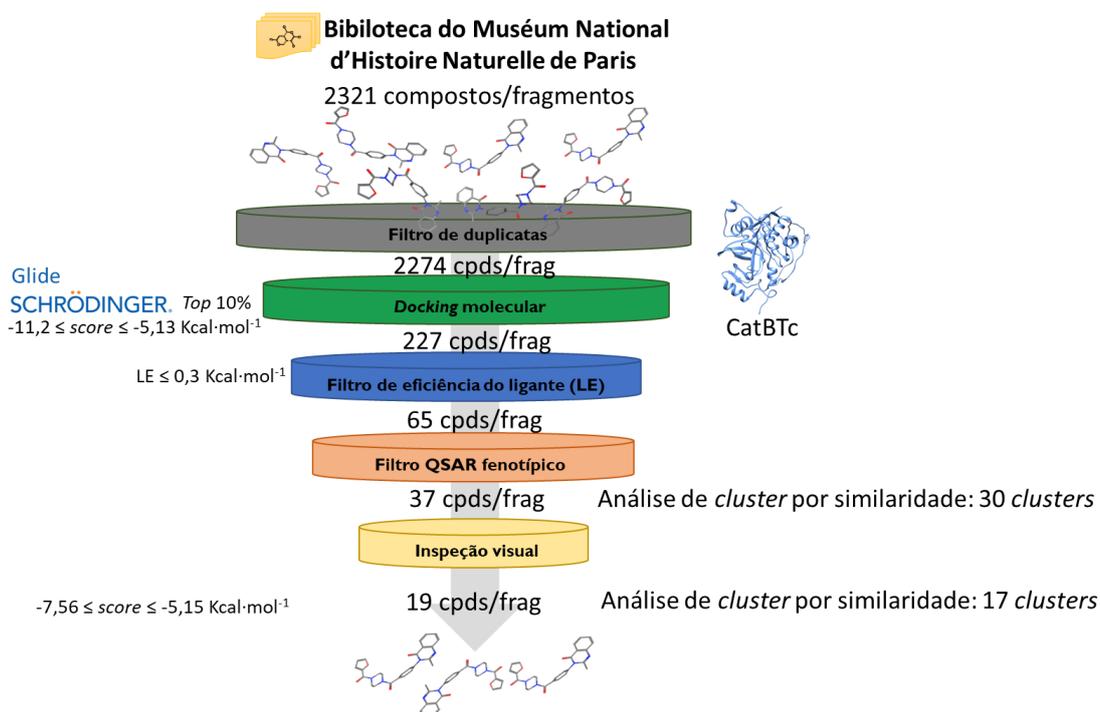
Após a validação das métricas, a biblioteca do MNHN, contendo 2321 compostos e fragmentos, passou pelo filtro de remoção de duplicatas, resultando em 2274 compostos/fragmentos. Posteriormente, os compostos foram submetidos aos cálculos de *docking* molecular e foram selecionados os *top* 10% compostos, ou seja, 227 primeiros. O *score* desses compostos variou de -11,2 à -5,13 Kcal.mol<sup>-1</sup>. Após o filtro de eficiência do ligante (LE), 65 compostos com LE acima de 0,3 Kcal.mol<sup>-1</sup> foram selecionados. Os compostos passaram, então, pelo filtro de QSAR fenotípico de *T. cruzi* (modelo com dados experimentais de 72h e 96h), sendo 37 compostos preditos como ativos (**Tabela 6**). Para esses 37 compostos, aplicamos análise de similaridade estrutural, através do programa DataWarrior (descriptor SkelSpheres, similaridade  $\leq 0,8$ ), que resultou em 30 *clusters* e 30 estruturas representativas.

**Tabela 6** - Compostos priorizados e seus respectivos scores de *docking*, eficiência do ligante (LE) e probabilidade de atividade frente a *T. cruzi* (predita pelo modelo QSAR fenotípico).

<b>Composto</b>	<b><i>Docking score</i> (Kcal·mol<sup>-1</sup>)</b>	<b>Eficiência do ligante (LE)</b>	<b>Probabilidade de atividade frente a <i>T. cruzi</i></b>
1153	-7,56	0,30	0,87
1612	-7,40	0,34	0,87
1696	-7,36	0,37	0,88
2024	-7,11	0,31	0,73
1597	-7,05	0,34	0,84
2292	-6,95	0,33	0,76
96	-6,77	0,31	0,82
730	-6,73	0,35	0,60
1139	-6,65	0,32	0,71
1123	-6,64	0,35	0,75
1138	-6,61	0,33	0,72
448	-6,47	0,32	0,74
732	-6,40	0,32	0,76
1594	-6,37	0,30	0,83
1699	-6,22	0,31	0,81
1685	-6,15	0,31	0,85
2151	-6,13	0,32	0,83
1694	-6,13	0,31	0,88
443	-6,11	0,47	0,65
1911	-6,01	0,30	0,84
726	-5,99	0,50	0,73
2137	-5,95	0,42	0,77
1673	-5,75	0,30	0,61
2271	-5,71	0,52	0,67
2157	-5,54	0,31	0,83
128	-5,52	0,31	0,60
1959	-5,38	0,38	0,66

<b>1668</b>	-5,36	0,30	0,82
<b>133</b>	-5,30	0,31	0,67
<b>2142</b>	-5,25	0,38	0,71
<b>530</b>	-5,23	0,31	0,69
<b>714</b>	-5,23	0,31	0,69
<b>739</b>	-5,20	0,32	0,67
<b>791</b>	-5,17	0,43	0,71
<b>2272</b>	-5,16	0,43	0,75
<b>1566</b>	-5,15	0,37	0,62
<b>426</b>	-5,15	0,37	0,71

Finalmente, a inspeção visual desses compostos foi efetuada e a partir das interações observadas e sobreposições com os PDBs da CatB *T.brucei* complexada ao ligante CA074 (3HHI), CatB *H. sapiens* complexada ao ligante nitroxolina (3AI8) e CatB *H. sapiens* complexada ao ligante Composto 28 (6AY2), 19 compostos foram selecionados e agrupados em 17 *clusters* apresentando *scores* de -7,56 à -5,15 Kcal.mol<sup>-1</sup> (**Figura 31**).



**Figura 31. Esquema do fluxograma da triagem virtual dos compostos provenientes do Museu Nacional de História Natural de Paris contra a CatB *T.cruzi*, utilizando diferentes abordagens computacionais.**

Os compostos e fragmentos 96, 448, 791, 1139, 1153, 1566, 1597, 1668, 1685, 1696, 1911, 1959, 2024, 2137, 2142, 2151, 2157, 2272 e 2292 foram os 19 priorizados após a triagem, mostrados na ordem de *score*. A escolha destes compostos e fragmentos ocorreu pela presença de interações com os resíduos do sítio catalítico (Cys120 e His278) e as histidinas pertencentes ao *loop* de oclusão (His192 e His193). A presença e interação do grupo carboxilato dos ligantes nas regiões próximas às histidinas do *loop* de oclusão foi identificado como um fator para a seleção dos compostos/fragmentos, assim como as interações diretas com os resíduos pertencentes à tríade catalítica. Observa-se que quase todos os compostos interagem com pelo menos um resíduo do *loop* de oclusão com exceção de 448, 1685 e 2142. Nota-se também a presença de resíduos extras fora dos indicados como importantes. Indicando regiões importantes na protease que são bem definidos na literatura. 13 compostos/fragmentos interagem com o resíduo Gln114 (1696, 2024, 1597, 96, 448, 1685, 1911, 2137, 2157, 1959, 1668, 2142 e 1566). 12 compostos interagem com o resíduo Gly277 (1153, 1696, 2024, 1597, 2292, 96, 1139, 448, 1685, 2151, 1959 e

2142). Outras interações aparecem com menor frequência: Ser115, Ser116, Cys117, Cys118, Glu205, Val255 e Trp300. A sobreposição da estrutura ou partes das estruturas dos compostos com os ligantes co-cristalizados provenientes dos PDBs, 3HHI, 3AI8 e 6AY2, foi também realizada nessa etapa indicando interações semelhantes (**Figura 32**).

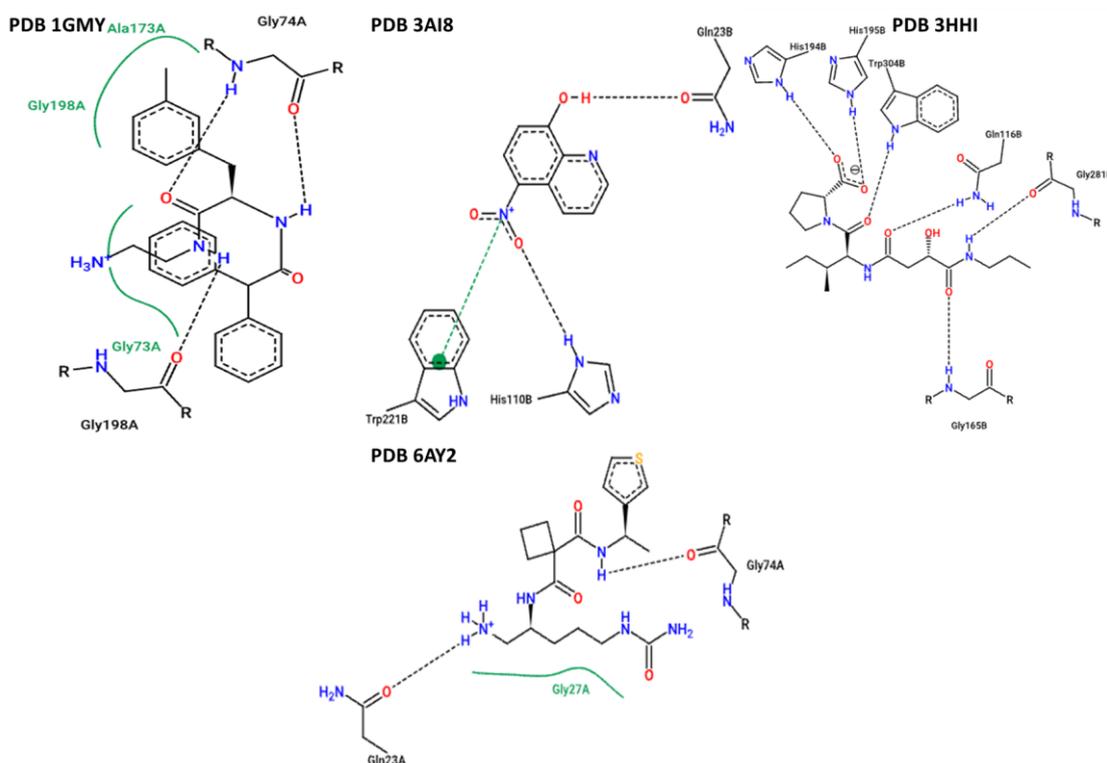
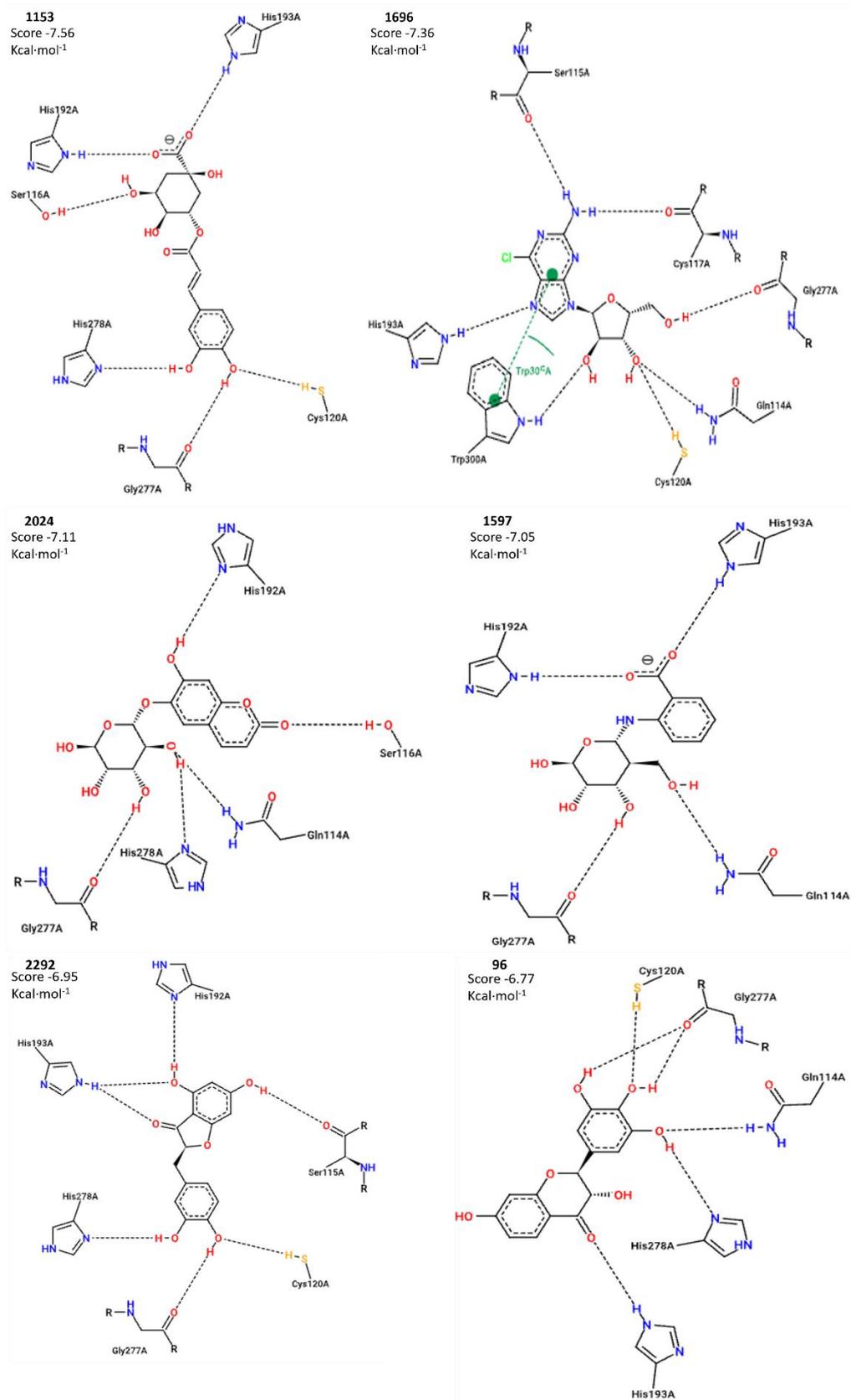
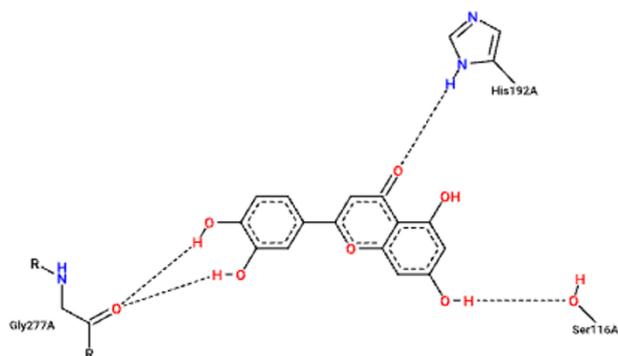


Figura 32. Esquema do diagrama 2D das interações dos ligantes co-cristalizados dos PDBs 1GMY (CatB *H. sapiens*), 3AI8 (CatB *H. sapiens*), 3HHI (CatB *T. brucei*) e 6AY2 (CatB *H. sapiens*), utilizados como comparação na inspeção visual dos resultados de *docking* da catepsina B *T. cruzi*.

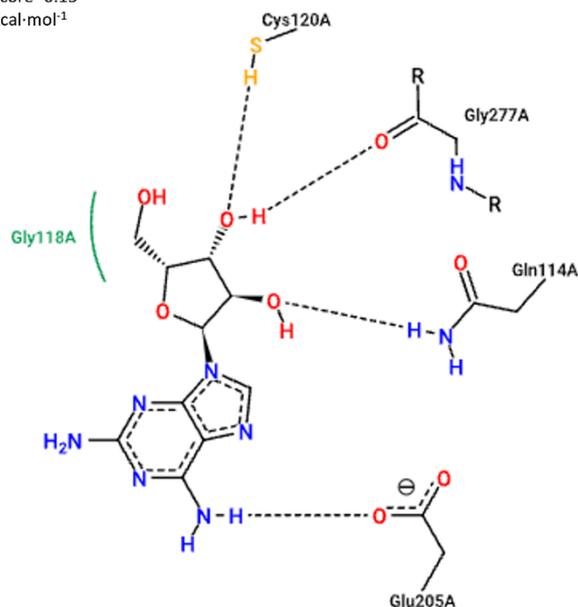


**Figura 33.** Esquema do diagrama 2D das interações dos **top 10** compostos e fragmentos selecionados com a CatBtc, apresentando os **6** primeiros. Figuras geradas pelo servidor ProteinsPlus, com a ferramenta PoseView e compostos mostrados de acordo com o *score* de *docking* em ordem crescente. Continua na Figura 32.

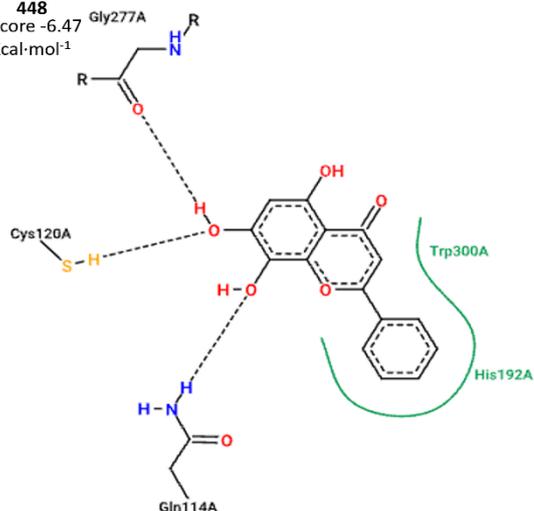
1139  
Score -6.65  
Kcal·mol<sup>-1</sup>



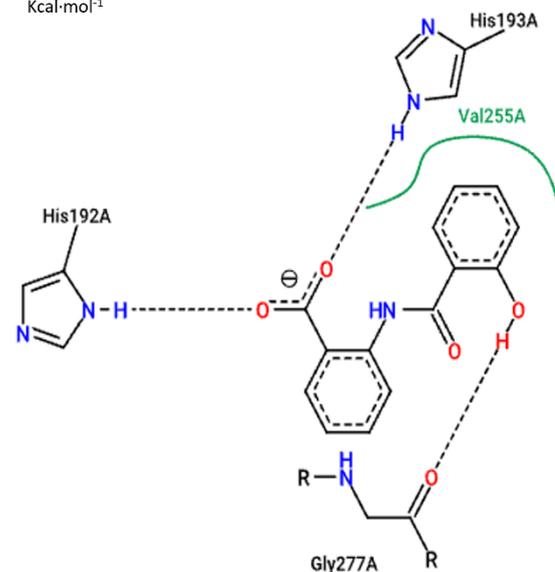
1685  
Score -6.15  
Kcal·mol<sup>-1</sup>



448  
Score -6.47  
Kcal·mol<sup>-1</sup>



2151  
Score -6.13  
Kcal·mol<sup>-1</sup>



**Figura 34. Conclusão do esquema do diagrama 2D das interações dos top 10 compostos e fragmentos selecionados com a CatBtc, apresentando os 4 últimos.** Figuras geradas pelo servidor ProteinsPlus, com a ferramenta PoseView e compostos mostrados de acordo com o score de docking em ordem crescente.

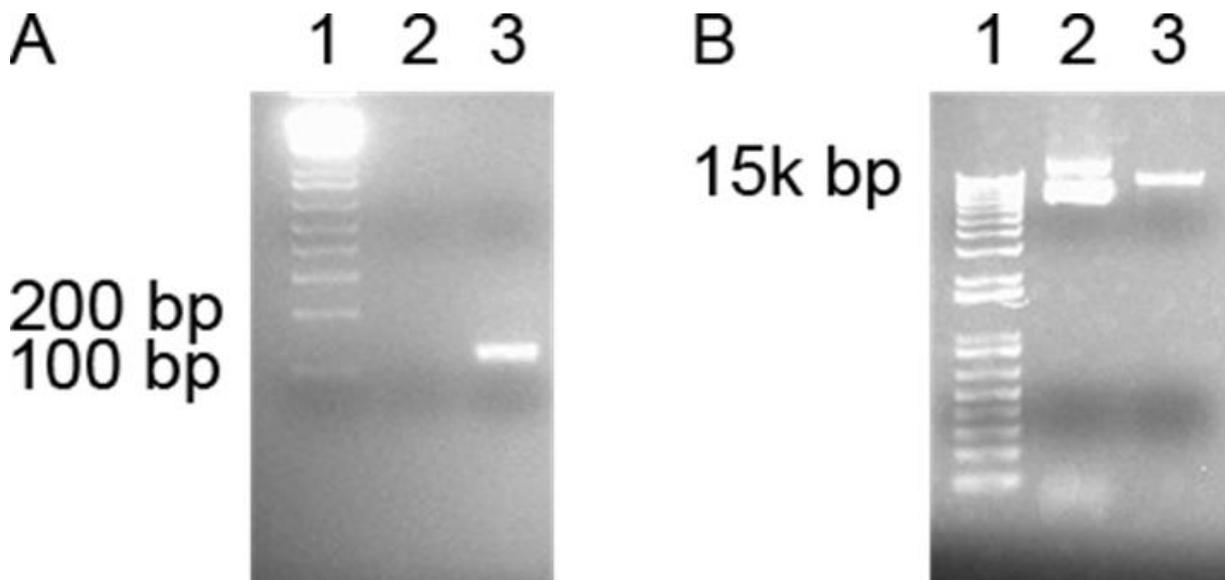
## PARTE II: DADOS EXPERIMENTAIS

### 5.5 Nocaute gênico e amplificação do sgRNA

Com o intuito de aprofundar o estudo da caracterização funcional da catepsina B de *T. cruzi*, optamos pelo nocaute gênico utilizando a técnica de CRISPR-Cas9.

Num primeiro momento, a região correspondente a sequência de sgRNA contendo 20 pb localizada nas posições 331-350 do gene foi escolhida, com o auxílio da ferramenta Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool (EuPaGDT), em virtude de sua conservação em ambos os alelos da cepa CL Brener. Os parâmetros considerados para sua escolha foram o *score* gerado pelo EuPaGDT, o conteúdo de GC, a localização na porção inicial do gene, indicando uma região flanqueadora propícia para o reparo por micro-homologia, a ausência de estruturas secundárias pelo programa RNAfold Web Server, como também a ausência de *off-targets*.

Então, o primeiro passo foi a amplificação, por PCR, da sequência de DNA correspondente ao sgRNA para que fosse posteriormente inserido no plasmídeo pTREX-n/Cas9. O *primer* senso contém a região específica do sgRNA e um sítio de restrição para *Bam*HI, já o *primer* anti-senso possui a sequência para a amplificação da estrutura secundária do sgRNA. Desse modo, o produto obtido da PCR foi de 122 pb (**Figura 35 A**). O plasmídeo pTREX-n/Cas9, próprio para a expressão constitutiva do sgRNA, foi digerido com a enzima de restrição *Bam*HI, assim como o sgRNA amplificado, para que ocorra a ligação. O plasmídeo possui 11,2k pb e como tem apenas dois sítios para a enzima de restrição, ele se encontrou linearizado (**Figura 35 B**).

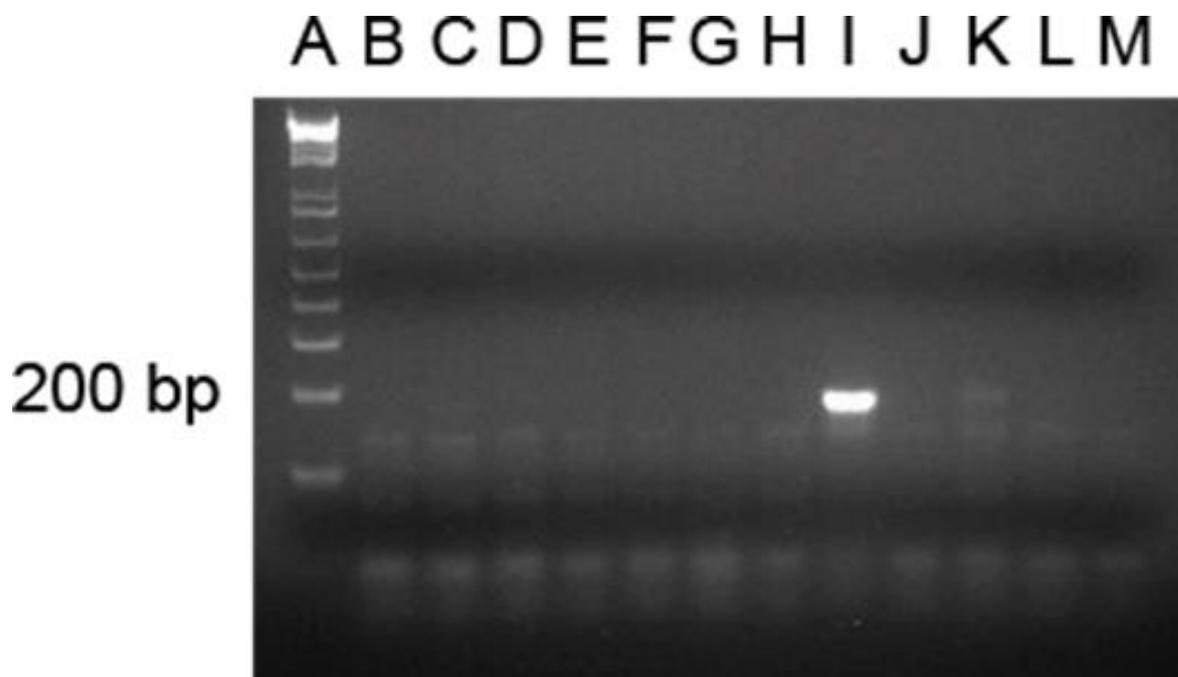


**Figura 35. A. Eletroforese em gel de agarose 2,5% da amplificação do sgRNA.** Marcador de peso molecular 1 Kb plus Invitrogen™ utilizado no poço 1; sgRNA no poço 3 com 122 pb. **B. Eletroforese em gel de agarose 1% da linearização do plasmídeo pTREX-n/Cas9.** Marcador de peso molecular

1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; poço 2, o plasmídeo antes da digestão e poço 3 com o plasmídeo digerido com *Bam*HI, ambos com 11,2k pb.

## 5.6 Preparação do plasmídeo pTREX-n/Cas9

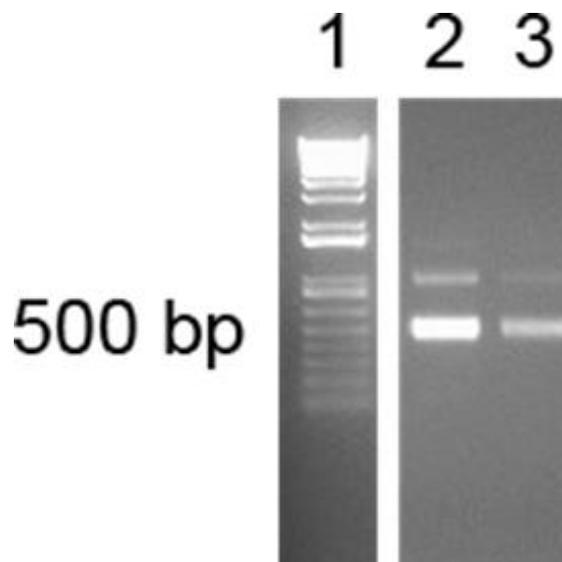
O plasmídeo pTREX-n/sgRNA/Cas9, após a ligação com o sgRNA, foi submetido a uma PCR de colônia para confirmação da correta orientação do sgRNA. O *primer* senso utilizado foi o mesmo que a amplificação do sgRNA e o *primer* anti-senso se anela na região correspondente ao gene *hx1* presente no plasmídeo pTREX-n/sgRNA/Cas9. Como o sgRNA foi clonado entre os sítios de restrição de *Bam*HI, a montante do gene *hx1*, apenas as colônias que estiverem orientadas corretamente amplificarão na PCR. Esta amplificação equivale a um inserto de 190 pb. Os poços foram nomeados com letras ordenadas alfabeticamente com colônias entre os poços B-L e o poço M como controle negativo. Como observado na **Figura 36**, apenas os poços I e K indicaram amplificação com o tamanho desejado, mas como o poço K estava com uma amplificação fraca, apenas o poço I foi selecionado.



**Figura 36. Eletroforese em gel de agarose 2,5% da confirmação da orientação correta do sgRNA clonado no pTREX-n/sgRNA/Cas9.** Os poços foram nomeados com letras ordenadas alfabeticamente de A-M. Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço A; as colônias estão distribuídas entre os poços B-L e o controle negativo está no poço M. sgRNA amplificado com tamanho de 190 pb nos poços I e K.

### 5.7 Amplificação do cassete de homologia

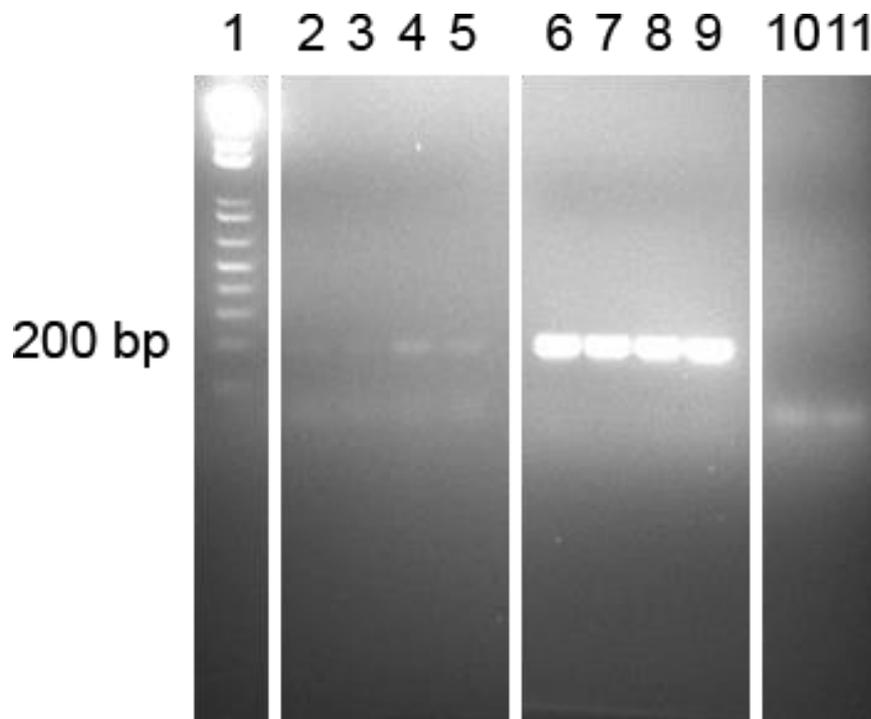
A estratégia utilizada para aproveitar do sistema de reparo por homologia direta do *T. cruzi*, no momento do corte do gene *catbtc*, foi introduzir um DNA doador para que fosse possível realizar a homologia com o gene e que ele continuasse truncado. Para tanto, a construção de um cassete contendo braços de homologia com o gene *catbtc* a montante e a jusante do gene da blasticidina (*bsd*) foi realizado para atuar como este DNA doador. Então, foram utilizados dois ultrâmeros para a amplificação do cassete, o ultrâmero senso contém 80 pb da região 5' UTR do gene *catbtc* logo antes do ATG seguidos de 20 pb do começo do gene *bsd* e o ultrâmero anti-senso possui os últimos 20 pb do gene *bsd* seguidos de 80 pb da região logo após o corte da *SpCas9* no gene *catbtc*. Dessa forma, após sua amplificação usando como molde o plasmídeo pGEM-BSD, o gene completo *bsd* será amplificado entre essas regiões do gene *catbtc*, resultando em um cassete de 556 pb. Na **Figura 37**, é possível observar o cassete amplificado com 556 pb. O poço 2 é correspondente a 2 $\mu$ l de cassete amplificado e o poço 3, a 1 $\mu$ l.



**Figura 37. Eletroforese em gel de agarose 1,5% da amplificação do cassete de homologia.** Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; poços 2 e 3 contendo diferentes concentrações do cassete amplificado com 556 pb. O poço 2 mostra a concentração correspondente a 2 $\mu$ l de cassete amplificado, enquanto o poço 3 corresponde a 1 $\mu$ l.

### 5.8 PCR de colônia de confirmação do cassete clonado

Após a amplificação do cassete, a clonagem no vetor pGEM®-T Easy (Promega™) foi realizada. Para a confirmação, uma PCR de colônia foi feita utilizando o sistema *White/blue*, ou seja, o plasmídeo possui uma região codificadora de  $\beta$ -galactosidase no sítio de múltipla clonagem que quando um gene é clonado nesta posição, sofre disrupção. Assim, o efeito disso é observado após a transformação em bactérias competentes, no plaqueamento em que são presenciadas colônias com as cores azul e branco. As colônias azuis mostram que o gene da  $\beta$ -galactosidase continua intacto e, conseqüentemente, não houve a clonagem, em alternativa, as colônias brancas indicam clonagem. Então, para a PCR, os *primers* senso e anti-senso usados amplificam uma região de 181 pb dentro do gene *bsd*. Na **Figura 38**, os poços 2-10 representam colônias da seguinte forma: 2-5 foram colônias brancas do cassete clonado no pGEM®-T Easy; 6-9, colônias brancas do plasmídeo pGEM\_BSD usado na amplificação do cassete, servindo de controle positivo; 10, colônia azul do cassete clonado no pGEM®-T Easy; 11, não foi utilizado nenhum DNA proveniente de controle funcionando com um controle negativo, juntamente com o poço 10, porém com um controle também para os *primers*. O poço 4 foi o escolhido para seguir com os experimentos por apresentar a melhor amplificação.



**Figura 38. Eletroforese em gel de agarose 2,5% da confirmação do cassete clonado no pGEM®-T Easy.** Marcador de peso molecular 1 Kb Plus DNA Ladder, Invitrogen™ utilizado no poço 1; O cassete amplificado possui 181 pb. Os poços 2-5 representam colônias brancas do cassete clonado no pGEM®-

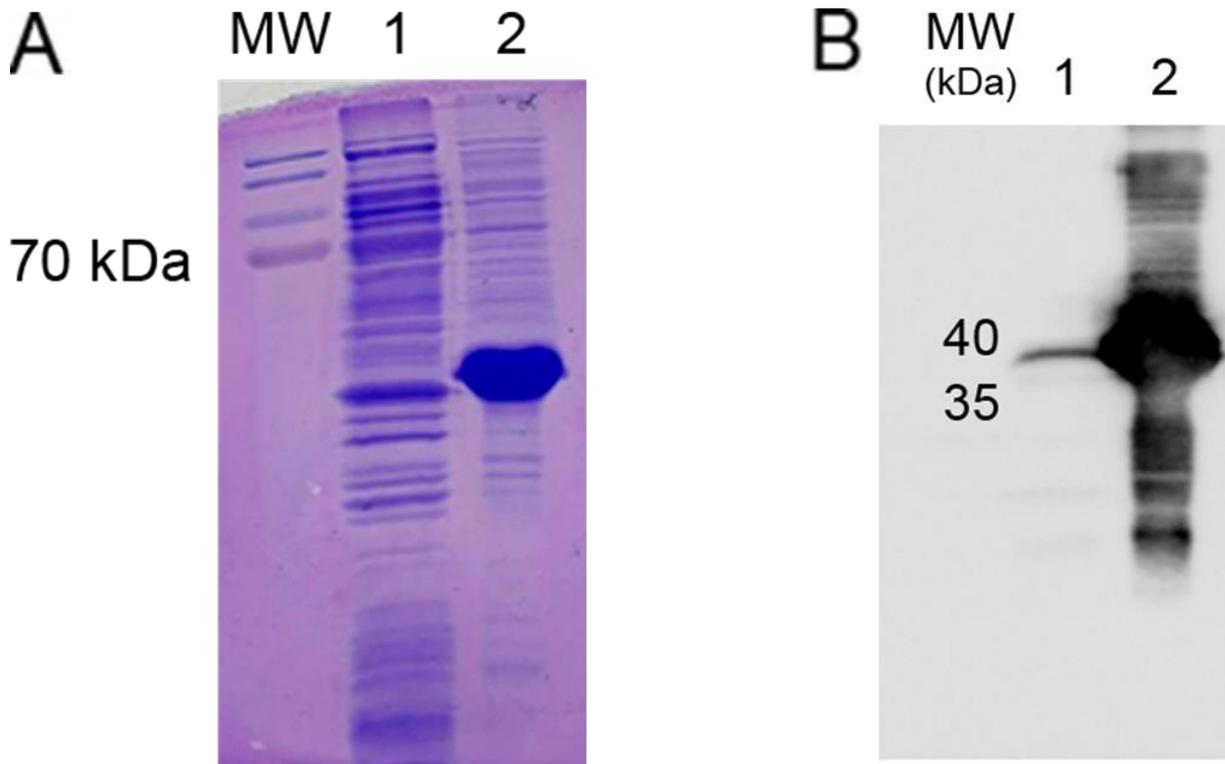
T Easy; os poços 6-9 são colônias brancas do pGEM-BSD; o poço 10 é uma colônia azul do cassete clonado no pGEM®-T Easy; o poço 11 não possui DNA, apenas os *primers*.

## 5.9 Expressão heteróloga da catepsina B de *Trypanosoma cruzi*

A obtenção da catepsina B de *T. cruzi* recombinante expressa de forma heteróloga em *E. coli* foi um dos objetivos para que se fizessem testes com a proteína. Para tanto, o plasmídeo pET-19B com a sequência da pró-catepsina B e a presença de uma *His-tag* na porção 3' da sequência foi clonada entre os sítios de *XhoI* e *BamHI* no sítio de múltipla clonagem. Este plasmídeo foi encomendado da empresa Genone.

### 5.9.1 Expressão em *E. coli* Rosetta (DE3) pLysS

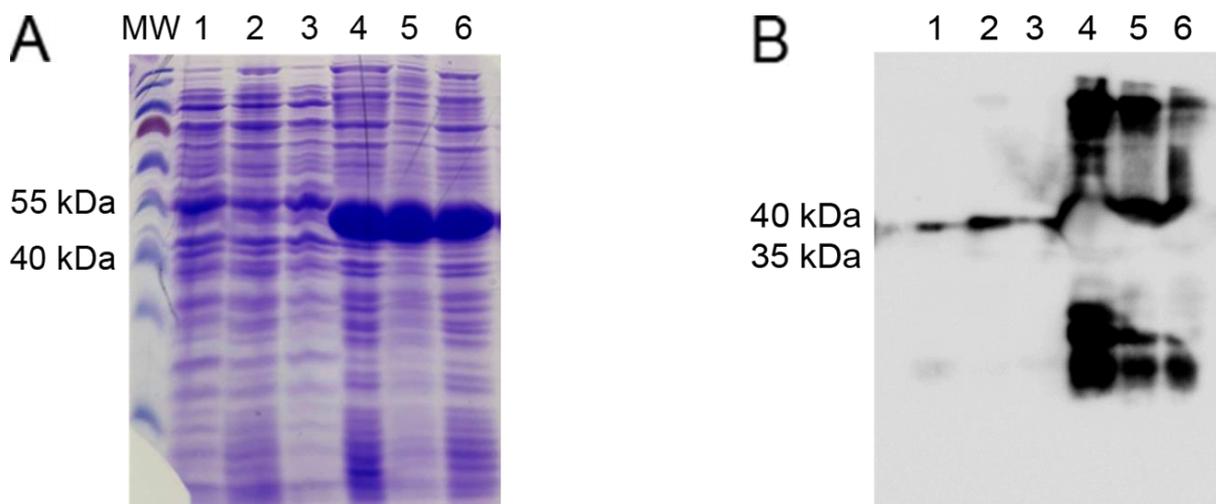
A fim de averiguar em qual fração a proteína seria expressa, a sua indução foi realizada em *E. coli* Rosetta (DE3) pLysS por 5 horas, 18°C, IPTG 0,5 mM e cloranfenicol na concentração de 34µg/ml. Como é possível observar na **Figura 39 A**, a pró-catepsina B foi expressa nas duas frações no peso molecular esperado de aproximadamente 37 kDa (fração solúvel no poço 1 e fração insolúvel no poço 2). Como também na **Figura 39 B**, foi detectada a presença da proteína nas duas frações por meio de *Western Blot*, porém em maior quantidade na fração insolúvel (fração solúvel no poço 1 e fração insolúvel no poço 2).



**Figura 39. Indução da pró-catepsina B com IPTG 0,5mM à 20°C, por 5 horas em *E. coli* Rosetta (DE3) pLysS. A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução da pró-catepsina B com o poço 1 com a fração solúvel, o poço 2 com a fração insolúvel e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da indução da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. O poço 1 representa a fração solúvel, o poço 2 a fração insolúvel e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™.

### 5.9.2 Expressão em *E. coli* BL21 (DE3)

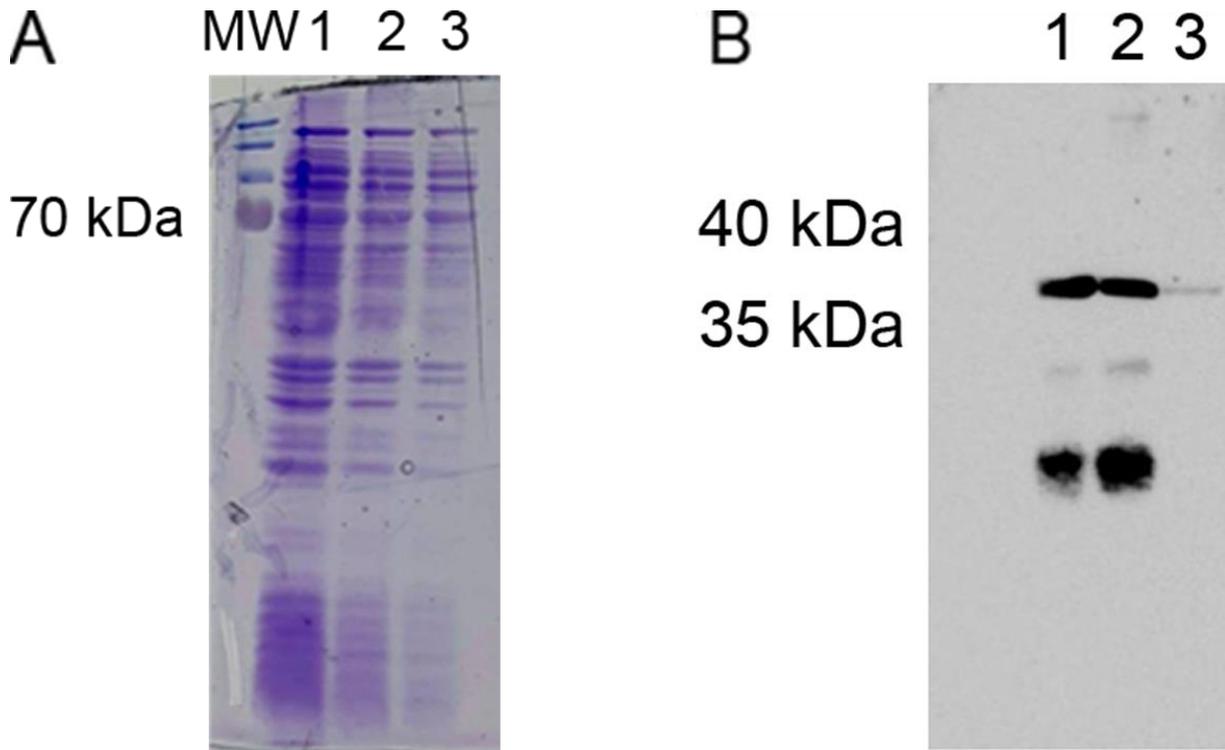
Para avaliar se era possível obter a proteína na fração solúvel, outra indução foi testada em *E. coli* BL21 (DE3). Os parâmetros utilizados foram: IPTG 1mM, por 3 horas a 37°C. Então, 6 amostras da indução foram analisadas (3 amostras de fração solúvel e 3 amostras de fração insolúvel). Na **Figura 40 A**, é possível observar a presença de bandas em todos os poços ainda com maior expressão na fração insolúvel. Na **Figura 40 B**, a proteína foi expressa em todas as frações no peso molecular esperado de aproximadamente 37 kDa com a confirmação da maior quantidade nas frações insolúveis. Em relação às frações solúveis, a expressão parece ter sido melhor na indução da cepa BL21 (DE3) em comparação à fração solúvel expressa na cepa Rosetta (DE3) pLysS.



**Figura 40. Indução da pró-catepsina B com IPTG 1mM à 37°C, por 3 horas em *E. coli* BL21 (DE3).** **A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução da pró-catepsina B com os poços 1-3 com frações solúveis, os poços 4-6 com frações insolúveis e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da indução da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. Os poços 1-3 representam as frações solúveis, os poços 4-6, as frações insolúveis e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™.

### 5.9.3 Análise das frações solúveis expressas

Para comparar a expressão das frações solúveis obtidas das diferentes induções, 2 amostras contendo a fração solúvel expressa em *E. coli* BL21 (DE3) e 1 amostra contendo a fração solúvel expressa em *E. coli* Rosetta (DE3) pLysS, foram submetidas ao gel NuPAGE™ Bis-Tris 4-12% e *Western Blot*. Nas **Figuras 41 A e B**, os poços 1 e 2 representam frações diferentes solúveis da indução prévia na cepa BL21 (DE3). O poço 3 representa a fração solúvel anteriormente induzida na cepa Rosetta (DE3) pLysS. Na **Figura 41 B**, a expressão na cepa BL21 (DE3) se mostra melhor quando comparada à realizada na cepa Rosetta (DE3) pLysS.

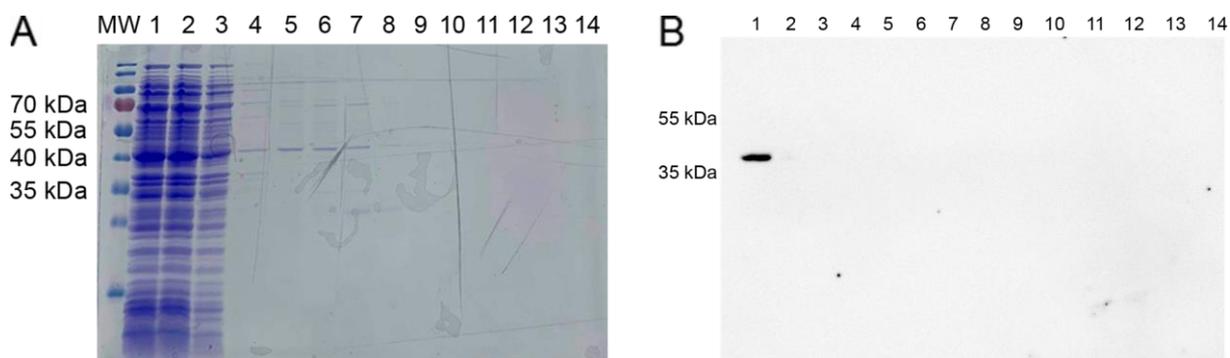


**Figura 41. Indução das frações solúveis previamente expressas da pró-catepsina B em *E. coli* cepas BL21 (DE3) e Rosetta (DE3) pLysS.** **A.** Gel SDS-PAGE 12% corado com azul de Coomassie da indução das frações solúveis da pró-catepsina B com os poços 1-2 correspondentes às expressões em BL21 (DE3); poço 3, em Rosetta (DE3) pLysS e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da indução das frações solúveis da pró-catepsina B com os poços 1-2 correspondentes às expressões em BL21 (DE3); poço 3, em Rosetta (DE3) pLysS e o MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™.

#### 5.9.4 Purificação da fração solúvel expressa em *E. coli* BL21 (DE3)

Como a fração solúvel expressa em *E. coli* BL21 (DE3) se mostrou a melhor porção comparada as outras induções feitas, a purificação com esta expressão foi realizada. As amostras de 500 µl cada foram purificadas com a resina Ni Sepharose™ High Performance e, então, avaliadas no gel NuPAGE™ Bis-Tris 4-12% (**Figura 42 A**) e, posteriormente, foi feito o *Western Blot* (**Figura 42 B**). O poço 1 corresponde ao extrato total; poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4; tampão de ligação final; poços 5-9, eluições com imidazol nas concentrações crescentes de 10mM, 20mM, 40mM, 60mM, 80mM, respectivamente; poços 10-14, eluições com imidazol na concentração de 250mM; MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. É possível observar na **Figura 42 A**, nos poços 4-8, a presença de uma banda única, porém no peso

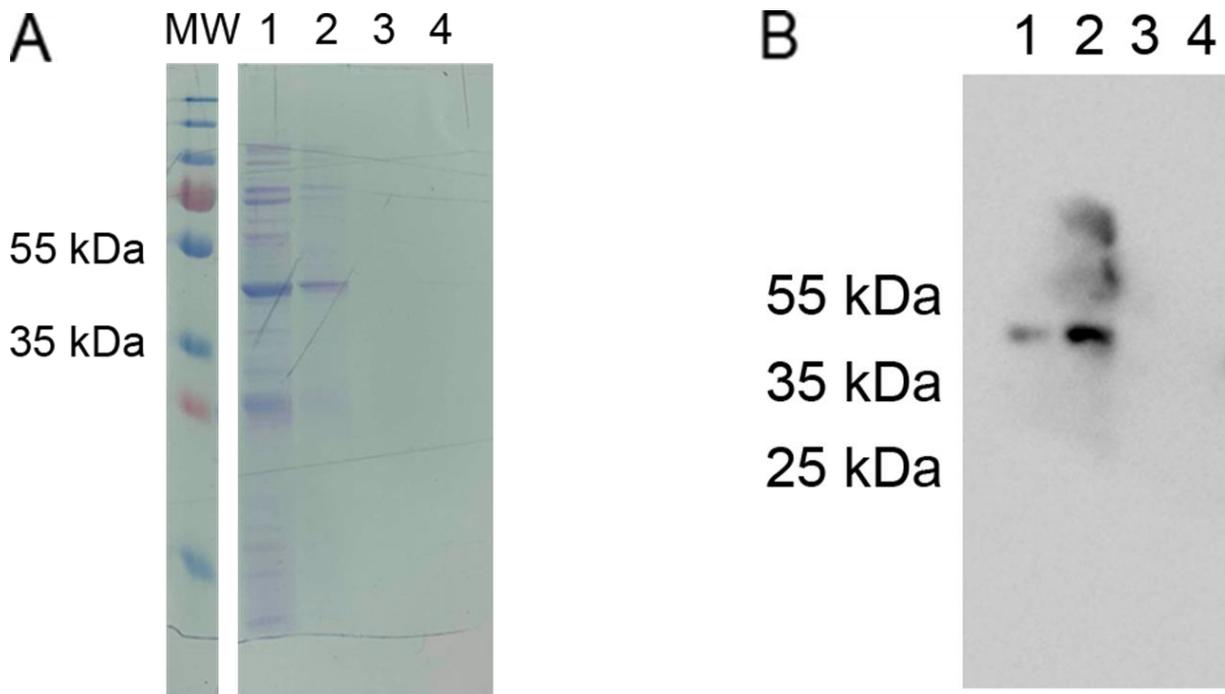
molecular de 40 kDa, um pouco acima do esperado de aproximadamente 37 kDa. Por outro lado, o resultado no *Western Blot*, indica apenas uma banda no peso molecular esperado no poço 1, correspondente ao extrato total, diferentemente do verificado no gel.



**Figura 42. Purificação da fração solúvel da pró-catepsina B a partir da expressão em *E. coli* BL21 (DE3).** **A.** Gel SDS-PAGE 12% corado com azul de Coomassie da purificação da pró-catepsina B com o poço 1 contendo o extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poços 5-9, eluições com imidazol 10mM, 20mM, 40mM, 60mM e 80mM, respectivamente; poços 10-14, eluições com imidazol 250mM e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da purificação da fração solúvel da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 contendo o extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poços 5-9, eluições com imidazol 10mM, 20mM, 40mM, 60mM e 80mM, respectivamente; poços 10-14, eluições com imidazol 250mM e MW, marcador de massa molecular PageRuler™ Prestained Protein Ladder, Thermo Scientific™.

### 5.9.5 Concentração da purificação da fração solúvel

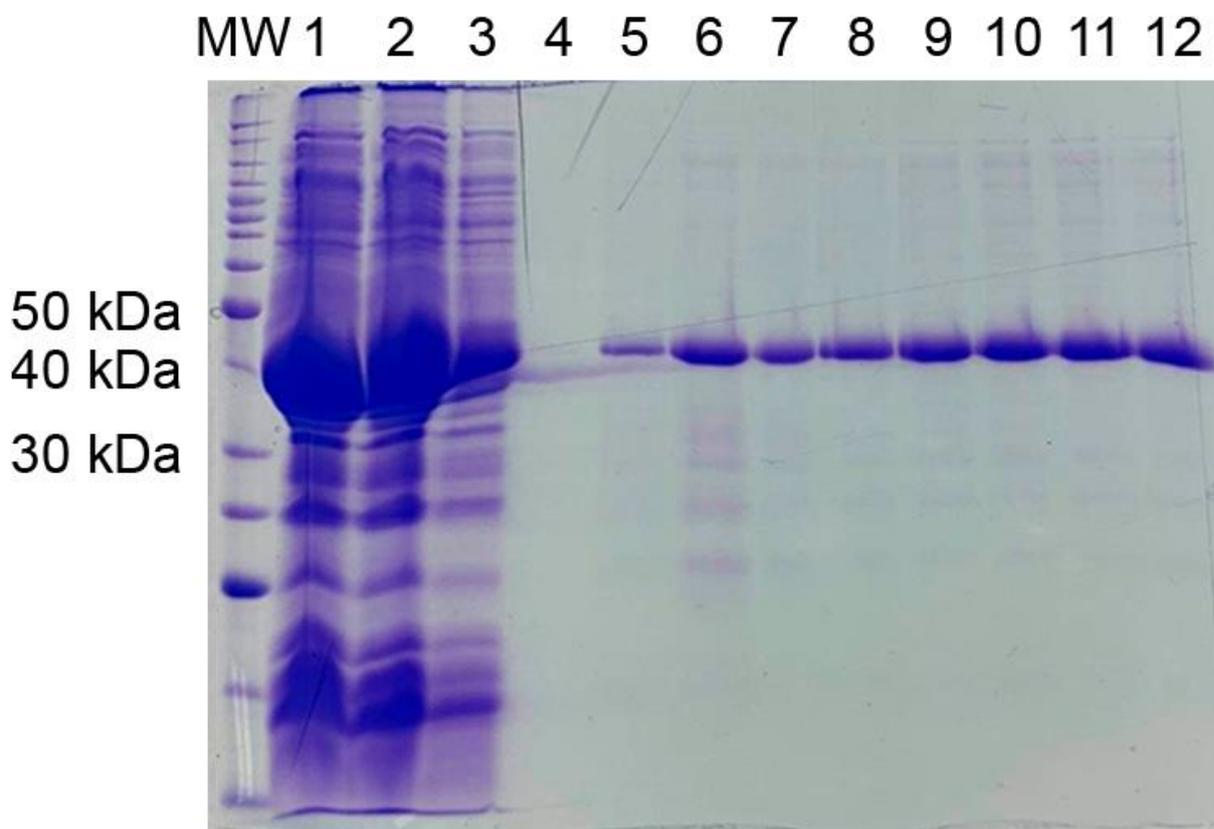
Foi levantada a hipótese de que devido às diversas eluições, a proteína tenha diluído bastante e, portanto, precisaria ser concentrada para poder avaliar a sua purificação. Então as amostras purificadas da fração solúvel foram concentradas em 2 amostras: uma correspondente às eluições com concentrações de imidazol entre 10mM-80mM (80-) e outra com as 5 eluições com concentração de imidazol de 250mM (80+). O lixo das concentrações, ou seja, as alíquotas que restam após as concentrações do 80- e 80+ foram igualmente submetidas aos testes no intuito de identificar se a proteína atravessou o filtro, indicados por L80- e L80+, respectivamente. Nas **Figuras 43 A e B**, o poço 1 representa o 80-; o poço 2, o 80+; o poço 3, o L80- e o poço 4, o L80+. Em ambas as figuras, observam-se bandas apenas nos poços 1 e 2 com maior expressão em 80+ (**Figura 43 B**).



**Figura 43. Concentração da purificação da fração solúvel da pró-catepsina B a partir da expressão em *E. coli* BL21 (DE3).** **A.** Gel SDS-PAGE 12% corado com azul de Coomassie da concentração da purificação da pró-catepsina B com o poço 1 contendo 80-; o poço 2, 80+; poço 3, L80-; poço4, L80 e MW, marcador de massa molecular PageRuler™ Plus Prestained Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da purificação da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 contendo 80-; o poço 2, 80+; poço 3, L80-; poço4, L80 e MW, marcador de massa molecular PageRuler™ Plus Prestained Protein Ladder, Thermo Scientific™.

#### 5.9.6 Purificação da fração insolúvel expressa em *E. coli* Rosetta (DE3) pLysS

A fração insolúvel correspondente à expressão na cepa Rosetta (DE3) pLysS foi igualmente purificada, porém com concentrações diferentes de imidazol. O poço 1 equivale ao extrato total; poço 2, ao não-ligado; poço 3, ao tampão de ligação inicial; o poço 4, ao tampão de ligação final; poço 5, eluição imidazol 10mM; poços 6-12, eluições imidazol 250mM. Verifica-se que a proteína purificou em todas as eluições, porém em maior quantidade nas eluições com imidazol 250mM (**Figura 44**).



**Figura 44. Purificação da fração insolúvel da pró-catepsina B a partir da expressão em *E. coli* Rosetta (DE3) pLysS. A.** Gel SDS-PAGE 12% corado com azul de Coomassie da purificação da pró-catepsina B com o poço 1 equivalente ao extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poço 5, eluição com imidazol 10mM; poços 6-12, eluições com imidazol 250mM e MW, marcador de massa molecular BenchMark™ Protein Ladder, Thermo Scientific™. **B.** *Western Blot* da purificação da fração insolúvel da pró-catepsina B utilizando o anticorpo anti-His na diluição 1:2.000. com o poço 1 equivalente ao extrato total; o poço 2, não-ligado; poço 3, tampão de ligação inicial; poço 4, tampão de ligação final; poço 5, eluição com imidazol 10mM; poços 6-12, eluições com imidazol 250mM e MW, marcador de massa molecular BenchMark™ Protein Ladder, Thermo Scientific™.

#### 5.9.7 Teste de atividade das porções purificadas

Obteve-se, então, a proteína purificada em ambas as frações solúvel e insolúvel expressas pelas cepas BL21 DE3 e Rosetta DE3 pLysS, respectivamente. A proteína foi ativada com a utilização do tampão acetato de sódio 2M pH 3,9 para alteração da amostra para o pH ótimo e suplementada com DTT com concentração final de 5mM. Porém, ao se realizar o teste de atividade com o substrato Z-Leu-Leu-Val-Tyr-AMC, não foi possível constatar nenhuma atividade em ambas as proteínas purificadas.

# DISCUSSÃO

## 6 DISCUSSÃO

A função de uma proteína está diretamente relacionada com a sua estrutura 3D nativa. Assim como estudar a sua conformação enovelada é fundamental para futura compreensão e prevenção de doenças e igualmente na aplicação de novos tratamentos para processos patológicos. Com o advento da tecnologia, a modelagem computacional tridimensional de proteínas se tornou uma metodologia extremamente utilizada pela alta precisão e redução de recursos envolvidos (WARDAH et al., 2019). A catepsina B, uma cisteína protease, já foi proposta como essencial no processo de virulência e sobrevivência de distintos organismos, principalmente tripanossomatídeos, assim como está presente em todas as formas do protozoário em *T. brucei* e *T. cruzi* (SIQUEIRA-NETO et al., 2018; GARCIA et al., 1998).

Através das análises *in silico* realizadas para a catepsina B de *T. cruzi* da cepa CL Brener foi possível notar que os alelos mesmo sendo provenientes de diferentes origens evolutivas, TcII e TcIII, apresentaram poucas divergências tanto na sequência de nucleotídeos quanto na sequência de aminoácidos codificados. Como esperado, essa conservação não se configura como 100%, entretanto, é alta, se estendendo às principais regiões como o *loop* de oclusão, sítio catalítico e cisteínas formadoras de ponte dissulfeto entre os alelos e até mesmo entre *T. brucei* e *H. sapiens*. Porém, percebe-se que a estrutura secundária proveniente de *H. sapiens* apresenta diferenças em relação a de *T. cruzi* gerada. Portanto, esta estrutura foi comparada ao modelo obtido pelo AlphaFold, que obteve alto grau de compatibilidade. A maioria das predições se encontravam nas mesmas regiões com poucos erros de um resíduo a mais ou menos, quando ocorreram. A predição inclusive dos primeiros aminoácidos com grande grau de desordem era esperada, pois se tratam do peptídeo sinal.

Além disso, o gene exibiu sintonia com os gêneros *Trypanosoma* e *Leishmania*, os quais possuem representantes importantes em infecções parasitárias humanas, como: *T. brucei*, *T. congolense*, *T. rangeli*, *T. vivax*, *L. infantum*, *L. major*, *L. mexicana* e *L. panamensis*. Dentro do gênero *Leishmania*, a catepsina B de diferentes espécies já foi descrita com funções distintas observadas através de testes enzimáticos (GERBABA; GEDAMU, 2013). Em *L. major*, a protease não apresenta atividade contra substratos proteicos conhecidos que geralmente são designados como ativos, por exemplo ao Z-Arg-Arg-AMC, contrariamente ao que ocorre em *L. mexicana* (SELZER

et al., 1997). Porém, mesmo assim ela ainda regula processos biológicos fundamentais. Em *L. mexicana*, foi mostrado *in vitro* que parasitos mutantes nulos para a peptidase reduziram a sobrevivência dentro dos macrófagos (BART et al., 1997). Assim como em *L. major*, que os parasitos nocautes apontaram uma resposta imune protetiva rápida com recrutamento celular do tipo Th1 (RASID et al., 2016; GONZALEZ-LEAL et al., 2014). No nocaute de *L. donovani* foi constatada a modulação diferencial das proteínas expressas com funções traducionais e ribossomais pelo parasito bem como redução na virulência (GERBABA; GEDAMU, 2013). No gênero *Trypanosoma*, em *T. brucei*, os mutantes nulos erradicaram a mortalidade nos testes *in vivo*, bem como reduziram as inflamações nos órgãos de camundongos infectados (ABDULLA et al., 2008). Todos esses resultados indicam o papel significativo da protease na interação parasito-hospedeiro em genes homólogos, e, conseqüentemente, a suposição da sua relevância em *T. cruzi*.

Na rede de interação predita de atuação da catepsina B em *H. sapiens* foi observado que a maioria dos genes pertencem e atuam no sistema imune. É descrita na literatura a relação da protease com as moléculas pertencentes ao MHC de classe II. Dentro dessa atuação, é visto que a catepsina B age junto à catepsina D na dissociação da cadeia invariável das moléculas de MHC de classe II (MIZUOCHI et al., 1994) e assim, denota a sua importância no processamento de antígenos. As proteínas COL15A1 e COL18A1 são responsáveis pela formação do tubo neuronal e estabilização de células musculares cardíacas. Já foi descrita a ação da catepsina B no aumento de endostatina, um inibidor endógeno da angiogênese produzido pela COL15A1 causando a má formação no tubo neuronal (KAZLAUSKAS, 2005). Os sistemas cardíaco e nervoso representam regiões conhecidas de atuação da protease. O papel na ativação por parte da catepsina B do inflamassoma NLRP3 é igualmente percebido na doença de Kawasaki, uma das principais doenças pediátricas cardíacas existentes. Por se caracterizar como um distúrbio altamente inflamatório, o tratamento é um desafio e, se não efetuado, aproximadamente 25% dos pacientes desenvolvem aneurisma coronário arterial. A probabilidade de tal quadro clínico ocorrer é aumentado na ativação do NLRP3 (JIA et al., 2019). Ademais, a pró-catepsina B já foi descrita na superfície celular de tumores, mas o seu papel só foi evidenciado com a descoberta da ligação com o tetrâmero de anexina II (ANXA2). A ANXA2 interage com a matriz celular, um indício da atividade de degradação da matriz extracelular seletiva durante invasão tumoral (MAI et al., 2000).

Em *T. cruzi*, através da predição de interação proteína-proteína, é possível observar a alta correlação na literatura com outras cisteíno proteases, principalmente a cruzipaina, a catepsina L em *T. cruzi*, visto que as proteases são descritas com funções e atuações próximas. Igualmente, é notada a relação com proteínas responsáveis pelas suas funções na membrana, como a sintaxina que participa na exocitose e a subunidade  $\beta$ -like da proteína ligante do nucleotídeo guanina que atua como homólogo da proteína G mandando sinais transmembrana. Proteínas de organização e compactação do Golgi, homólogas da GRASP humana aparecem também ressaltando a importância dos processos relacionados à exportação e atuação na membrana da peptidase no parasito, processos os quais ocorrem preferencialmente nas formas amastigotas do *T. cruzi* (ATWOOD et al., 2005).

Com objetivo de explorar sua conformação e suas funções, a modelagem da estrutura tridimensional da catepsina B de *T. cruzi* foi realizada. O modelo permitiu a inspeção da predição dos 3 sítios de N-glicosilação. O resíduo Asn74, se encontra na porção pró-catepsina e é clivada na maturação da enzima. O resíduo Asn195 predito faz parte do *loop* de oclusão e está conservado com o resíduo Asn113 da catepsina B de *H. sapiens* (CTSB), igualmente encontrado na região do *loop*, mas não se encontra conservado ao ser comparado com o PDB 4HWY da pró-catepsina de *T. brucei*. O resíduo Asn212 não está conservado com a CTSB, porém está com a Asn216 de *T. brucei*, que é apontado como sítio de glicosilação na espécie (REDECKE et al., 2013). Apesar da conservação dos resíduos entre a CatBTc e CTSB, esta asparagina não está descrita na literatura como sítio de glicosilação em *H. sapiens*, e sim o resíduo Asn289, conservado no modelo em *T. cruzi*, porém não é predito pelo programa (CHAN et al., 1986). Estes dados evidenciam as diferenças entre as catepsinas B provenientes de diferentes organismos, porém a manutenção da glicosilação é necessária, visto que, no aparato de Golgi ocorre a interação dos resíduos glicosilados com os receptores de manose-6-fosfato, viabilizando a maturação da protease com a sua entrada, através de vesículas, nos lisossomos (MACH et al., 1992). Na O-glicosilação, que ocorre nos resíduos de serina (Ser) e treonina (Thr), foi visto que dos 14 sítios identificados, 6 são encontrados na porção pró-catepsina e são clivados e 6 estão presentes no *loop* de oclusão (Ser189, Ser196, Ser197, Ser200, Ser203 e Thr208).

No tangente aos cálculos de *docking* realizados, os modelos trRosetta e AlphaFold foram usados na etapa de validação do protocolo de *docking*, visto que se

mostraram os melhores modelos dentre os gerados, além de utilizarem a mesma metodologia de modelagem por aprendizado de máquina. A fim de que não houvesse viés por parte dos resultados de apenas um programa de *docking*, o DockThor e o Glide foram aplicados para observar como os modelos se comportavam ligados aos compostos selecionados.

A catepsina B possui atividade dupla apresentando sua atividade ótima de exopeptidase em pH ácido de 4,5. A atividade de endopeptidase, mesmo que menor, ocorre em pH neutro, onde o *loop* de oclusão assume uma nova conformação mais aberta expondo o sítio ativo para compostos maiores (NÄGLER et al., 1997). Uma alta similaridade estrutural foi observada ao comparar os modelos obtidos pelos programas trRosetta e AlphaFold sobrepostos, porém a região do *loop* entre as duas estruturas representava as maiores diferenças de conformação. Como os modelos poderiam representar conformações correspondentes a diferentes atividades da protease e as validações dos *dockings* (com ativos e *decoys*) foram efetuadas tanto em pH 4,5 e 7,4. Assim, foi possível observar que, infelizmente, a validação de *docking* com ambos os modelos de CatBTc no programa DockThor não apresentou boas métricas, em nenhum dos pHs. A validação de *docking* utilizando o programa Glide na versão XP, mostrou melhores métricas para o modelo AlphaFold (em pH 4,5), em relação aos *top* 10% e esse protocolo foi selecionado para a triagem dos compostos MNHN.

Após as diversas etapas para seleção de compostos, observando os melhores *scores* de *docking*; eficiência de ligante  $> 0,3$ ; probabilidade de atividade frente a *T. cruzi*  $> 0,6$ ; e inspeção visual criteriosa das poses de *docking*, 19 compostos/fragmentos foram selecionados.

Os compostos e fragmentos 1153, 1696, 2292, 96, 448, 1685 e 2142 foram priorizados conforme presença de interação com resíduos importantes e/ou sobreposição com os ligantes co-cristalizados dos PDBs utilizados. Esses compostos apresentaram necessariamente interação com a Cys120 e ao menos uma ou várias interações com os resíduos His192, His193 e His278, indicando alta interação e encaixe no sítio catalítico e *loop* de oclusão da catepsina B de *T. cruzi*. Já os compostos 2024, 1597, 2292, 1139, 2151, 1911, 2137, 2157, 1959, 1668, 791, 2272 e 1566 mostraram sobreposição, especialmente com o PDB 3HHI, cujos grupamentos COO/NOO ocuparam a mesma posição no ligante co-cristalizado CA-074 dispondo das mesmas interações na região do sítio catalítico da proteína. O PDB 3AI8 possui

um pequeno ligante aromático que, apesar de se assemelhar com diversos fragmentos da lista, não se sobrepõe a nenhum composto ou fragmento. Em relação ao PDB 6AY2, o ligante cristalizado apesar de ser grande, não ocupa o mesmo espaço que os ligantes dos outros PDBs e os compostos testados. Isso mostra que mesmo possuindo um sítio catalítico altamente conservado entre diferentes espécies, a catepsina B de *T. cruzi* apresenta diferentes regiões de ligação com a evidência de um largo sítio catalítico. Ainda, constata-se que há a presença de fragmentos com menos de 300 kDa na lista e mesmo com pesos moleculares e tamanhos diferentes dos compostos, apontam capacidade ótima de ligação e interação com os resíduos catalíticos, vide resultados do filtro de eficiência do ligante.

A catepsina B de *T. cruzi* foi caracterizada pelo grupo e vem sendo estudada há alguns anos, porém sua funcionalidade ainda não havia sido explorada. Tendo em vista a análise funcional da protease para se adquirir um melhor entendimento da sua atuação na infecção, a metodologia de CRISPR/Cas9 foi utilizada para gerar um nocaute duplo no gene *catbtc*. Apesar do RNAi ser uma das técnicas mais utilizadas em tripanossomatídeos, visto que a técnica se mostra extremamente eficaz em *T. brucei*, por exemplo, o mecanismo é inexistente em *T. cruzi*. Assim, é necessário o uso de outro tipo de nocaute na espécie (BALAÑA-FOUCE; REGUERA, 2007; NGÔ et al., 1998; ULLU et al., 2002).

Para tal, a clonagem de um sgRNA, proveniente de uma sequência presente nas duas cópias do gene, no plasmídeo Cas9/pTREX-n foi obtido. Este plasmídeo eficientemente provoca a edição gênica de genes endógenos de *T. cruzi*, com a atuação da SpCas9 no gene de interesse, sem a detecção da toxicidade, comumente notada, da endonuclease. Além disso, propicia a expressão constitutiva tanto da endonuclease quanto do sgRNA, formando parasitos selecionados fluorescentes, pela presença do GFP, em uma única rodada de transfecção (DOCAMPO, 2015). Com o intuito de aumentar a eficiência da técnica, a construção de um cassete para a indução da recombinação homóloga foi igualmente efetuada. A estratégia se baseia na inserção de um cassete contendo o gene *bsd* flanqueado por uma sequência pertencente ao 5' UTR da *catbtc* e a sequência posterior ao corte do sgRNA da *catbtc* dentro do gene da *catbtc* ocasionando uma disrupção do mesmo com a recombinação homóloga das regiões provenientes da *catbtc*, assim, introduzindo o gene *bsd* presente no cassete (**Figura 13**).

Porém, após três tentativas de nocaute nas cepas CL Brener e Y de *T. cruzi*, não foi possível a obtenção de parasitos selecionados. Mesmo com a utilização do plasmídeo sgRNA/Cas9/pTREX-n conjuntamente com o cassete, os parasitos controle transfectados selecionados com os antibióticos neomicina e blasticidina não morreram sobre a pressão seletiva aplicada, porém os transfectados não suportaram a seleção e morreram. Uma das primeiras explicações poderia ser uma resistência criada aos antibióticos, o que esclareceria o resultado observado no controle. O *T. cruzi* é descrito como um parasito com alta capacidade de troca genética, e, portanto, da mesma forma apresenta capacidade de adquirir resistência aos antibióticos neomicina e higromicina, por exemplo (MILES, 1999). Outro ponto importante é que as deleções com êxito descritas na literatura não afetaram genes essenciais, os chamados *housekeeping*, ao parasito tal como: o gene codificador de um transportador mitocondrial de cálcio, TcMCU (CHIURILLO et al., 2017) e os genes TcPFR1 e TcPRF2, codificadores da proteínas paraflagelares 1 e 2, respectivamente, formadoras da estrutura da haste paraflagelar conectada aos microtúbulos do flagelo do parasito (DOCAMPO, 2015). Trabalhos que envolveram o nocaute de genes que estão envolvidos na captação de ácidos graxos,  $\beta$ -oxidação (*ech*) e síntese de DNA (*dhfr-ts*), por exemplo, considerados essenciais ao *T. cruzi*, resultaram na morte da população transfectada (XU et al., 2009; PENG et al., 2015). Um grande indício de que a catepsina B possa representar um gene essencial à sobrevivência do *T. cruzi*.

No que diz respeito à expressão heteróloga da proteína, a catepsina B para se tornar ativa, precisa ser expressa com a região pró, pois esta possui um mecanismo de enovelamento assistido por pró-sequência, o qual está presente em múltiplas proteases. Este mecanismo consiste na impossibilidade da proteína desnovelada formar sua estrutura tridimensional biologicamente ativa sem a sua região pró. O recurso abrange inclusive proteínas maduras desnaturadas *in vitro* e proteínas expressas em sistemas artificiais. A ativação da protease pode ocorrer ainda que a pró-sequência seja inserida após uma proteína estar desnovelada, se a ligação entre elas se caracteriza como não-covalente. (EDER; FERSHT, 1995; VAN DEN HAZEL; KIELLAND-BRANDT; WINTHER, 1993). À vista disso, a clonagem da pró-catepsina se torna necessária para a sua correta expressão. Apesar de se obter a proteína expressa tanto na fração solúvel quanto nos corpos de inclusão, nenhuma das induções resultou em uma proteína ativa ainda que testada em diferentes parâmetros e pHs. A catepsina B de *T. cruzi* dispõe de histórico descrito na literatura da dificuldade

de ser expressa ativa em sistema heterólogo. Em trabalhos anteriores realizados pelo grupo, a protease foi expressa em baculovírus, *Komagataella phaffii*, anteriormente conhecida como *Pichia pastoris* e *E. coli* (NÓBREGA, 2001; RESTREPO, 2005). Todas as expressões foram feitas com diferentes vetores e estratégias, porém não resultaram em uma proteína recombinante ativa. Ainda, a peptidase foi obtida apenas em corpos de inclusão e não foi expressa na sua fração solúvel. Um ponto importante notado nas tentativas anteriores que influenciou a expressão neste trabalho foi a clonagem do gene codificando a proteína completa ou fusionada a outras proteínas. A escolha da expressão em *E. coli* Rosetta (DE3) pLysS foi devido às mutações presentes na cepa que codificam seis tRNAs raros para a bactéria que propiciam uma melhora na tradução devido aos códons diferenciados gerados. A insuficiência de tRNAs pode gerar problemas na tradução como: terminação prematura, parada do processo e incorporação de aminoácidos errados. A cepa ainda produz a T7 lisozima, um inibidor natural da T7 RNA polimerase, suprimindo sua expressão basal. Assim, estabilizando recombinantes no vetor que codificam proteínas que afetam o crescimento e a viabilidade celular. Dessa forma, esta modificação favorece a expressão de proteínas heterólogas (KANE, 1995; NOVY et al., 2001; JOSEPH; PICHAIMUTHU; SRIMEENAKSHI, 2015). Visto que a expressão na cepa Rosetta (DE3) pLysS ou outra que tenha capacidade de gerar modificações no citoplasma ainda não havia sido testada, a cepa foi a primeira escolha para a expressão (NÓBREGA, 2001; RESTREPO, 2005). Sob pretexto da expressão heteróloga nas cepas empregadas ser viável na fração solúvel, a indução em outras cepas, como Origami B (DE3) e BL21-AI, pode resultar em uma proteína ativa. A cepa Origami B (DE3) carrega as características desejadas das cepas BL21 e Origami com capacidade de aumentar a formação de pontes de dissulfeto no citoplasma, assim minimizando a redução no citoplasma, o que facilita o correto enovelamento de proteínas (XIONG et al., 2005; BESSETTE et al., 1999). A cepa BL21-AI é capaz de expressar genes que podem ser tóxicos em outras cepas derivadas de BL21, além do T7 RNA polimerase ser mais estritamente regulado em comparação à BL21 (DE3), apresentando uma expressão basal maior do promotor (INVITROGEN, 2006). Novas tentativas e ajustes de indução na presença ou não de aditivos, devem ser realizadas a fim de esgotar as condições para obtenção da enzima ativa que não foram possíveis devido as dificuldades impostas pela pandemia da COVID-19.

# CONCLUSÃO

## 7 CONCLUSÃO

Os resultados *in silico* obtidos neste trabalho possibilitaram uma melhor compreensão da catepsina B de *T. cruzi* em relação a suas propriedades físico-químicas, estrutura tridimensional, conformações e sítio de ligação a ligantes. Além disso, a integração de diferentes abordagens computacionais possibilitou a priorização de compostos potenciais que apresentaram afinidade em relação a CatBTc, interagindo no seu sítio catalítico, além de terem sido preditos como ativos frente ao *T. cruzi*, nos modelos de QSAR fenotípico.

Também, foi possível a obtenção da proteína expressa e purificada em *E. coli* na fração solúvel como também nos corpos de inclusão. O nocaute gênico com a utilização do sistema CRISPR/Cas9 não foi alcançado nos parasitos sugerindo que a catepsina B de *T. cruzi* se caracteriza como um gene essencial ao parasito. No entanto, novas tentativas devem ser realizadas a fim de ratificar o envolvimento da catepsina B em processos importantes e seu potencial como futuro alvo de fármacos para o tratamento da doença de Chagas.

# **PERSPECTIVAS**

## 8 PERSPECTIVAS

As principais perspectivas deste trabalho são:

- A indução da catepsina B de *T. cruzi* heteróloga em outras cepas de *E. coli*, como por exemplo, Origami e BL21-AI, com o fim de avaliar a sua expressão e atividade;
- A obtenção de uma proteína recombinante ativa (rCatBTc);
- Produção de anticorpo a partir da proteína recombinante (rCatBTc);
- Repetição dos testes de nocaute gênico de CRISPR/Cas9 com o objetivo de se conseguir uma cultura mutante selecionada;
- Análises por meio de PCR, *Western Blot* e/ou imunofluorescência dos alelos deletados e expressão proteica diferencial do parasito mutante;
- Testes *in vitro* de inibição da atividade tanto com o extrato do parasito quanto com a proteína recombinante com todos os compostos promissores;
- Identificação de compostos inibitórios seletivos da catepsina B de *T. cruzi*.

# **REFERÊNCIAS BIBLIOGRÁFICAS**

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABDULLA, M. et al. RNA interference of *Trypanosoma brucei* cathepsin B and L affects disease progression in a mouse model. **PLoS Neglected Tropical Diseases**, v. 2, n. 9, p. 1–6, 2008.
- ACHARYA, K. R.; LLOYD, M. D. The advantages and limitations of protein crystal structures. **Trends in Pharmacological Sciences**, v. 26, n. 1, p. 10–14, 2005.
- ALMAGRO ARMENTEROS, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. **Nature Biotechnology**, v. 37, n. 4, p. 420–423, 2019.
- ALVES, V. M. et al. Cheminformatics: An introduction. **Quimica Nova**, v. 41, n. 2, p. 202–212, 2018.
- ANISIMOVA, M.; GASCUEL, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. **Systematic Biology**, v. 55, n. 4, p. 539–552, 2006.
- ATWOOD, J. A. et al. Microbiology: The *Trypanosoma cruzi* proteome. **Science**, v. 309, n. 5733, p. 473–476, 2005.
- BALANÍA-FOUCE, R.; REGUERA, R. M. RNA interference in *Trypanosoma brucei*: a high-throughput engine for functional genomics in trypanosomatids? **Trends in Parasitology**, v. 23, n. 8, p. 348–351, 2007.
- BARRANGOU, R.; HORVATH, P. A decade of discovery: CRISPR functions and applications. **Nature Microbiology**, v. 2, n. June, p. 1–9, 2017.
- BARRETT, A. J.; RAWLINGS, N. D. **Families and clans of cysteine peptidases** *Perspectives in Drug Discovery and Design*, 1996.
- BART, G. et al. Cathepsin B-like cysteine proteinase-deficient mutants of *Leishmania mexicana*. **Molecular and Biochemical Parasitology**, v. 88, n. 1–2, p. 53–61, 1997.
- BENKERT, P.; TOSATTO, S. C. E.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. **Proteins: Structure, Function and Genetics**, v. 71, n. 1, p. 261–277, 2008.
- BERN, C. Antitrypanosomal Therapy for Chronic Chagas' Disease. **New England Journal of Medicine**, v. 365, n. 13, p. 1258–1259, 2011.
- BERTHOLD, M. R. et al. KNIME: The Konstanz information miner. **4th International Industrial Simulation Conference 2006, ISC 2006**, v. 11, n. 1, p. 58–61, 2006.
- BERTONI, M. et al. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. **Scientific Reports**, v. 7, n. 1, p. 1–15, 2017.
- BESSETTE, P. H. et al. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 24, p. 13703–13708, 1999.
- BOND, J. S. **Proteases: History, discovery, and roles in health and disease** *Journal of Biological Chemistry*, 2019.
- BRASIL. Doença de Chagas aguda no Brasil: série histórica de 2000 a 2013. **Boletim Epidemiológico**, v. 46, n. 21, 2015.
- BRENER, Z. *Trypanosoma cruzi*: 1997.

BRINGAUD; RIVIÈRE; COUSTOU, 2006. Energy metabolism of trypanosomatids : Adaptation to available carbon sources. v. 149, p. 1–9, 2006.

BRYAN T. MOTT, RAFAELA S. FERREIRA, A. S. A. J. K. K.-H. A. L. M. S. J. T. S. P. S. D. M. R. A. J. H. M. J. I. C. P. A. C. J. T. B. K. S. AND D. J. M. Identification and optimization of inhibitors of Trypanosomal cysteine proteases: Cruzain, Rhodesain and TbCatB. **J Med Chem.**, v. 14, n. 53(1), p. 52–60, 2010.

BUSCHINI. Genotoxicity Revaluation of Three Commercial Nitroheterocyclic Drugs: Nifurtimox, Benznidazole and Metronidazole. **Journal of Parasitology Research**, 2009.

C. ONYEKWELE, K. Life Cycle of Trypanosoma cruzi in the Invertebrate and the Vertebrate Hosts . **Biology of Trypanosoma cruzi** , 2019.

CAVALLO-MEDVED; MOIN; SLOANE, 2011. Cathepsin B. **Physiology & behavior**, v. 176, n. 3, p. 139–148, 2011.

CHAGAS, C. Nova tripanozomíaze humana. **Memórias do Instituto Oswaldo Cruz**, v. 1, p. 159–218, 1909.

CHAN, S. J. et al. Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 20, p. 7721–7725, 1986.

CHIRAC, P.; TORREELE, E. Global framework on essential health R&D. **Lancet**, v. 367, n. 9522, p. 1560–1561, 2006.

CHIURILLO, M. A. et al. Different Roles of Mitochondrial Calcium Uniporter Complex Subunits in Growth and Infectivity of Trypanosoma cruzi. **mBio**, v. 8, n. 3, p. 1–16, 2017.

CHO, A. News flash: X-ray laser produces first protein structure. **Science**, v. 338, n. 6111, p. 1136, 2012.

CIANNI, L. et al. Design, synthesis and stepwise optimization of nitrile-based inhibitors of cathepsins B and L. **Bioorganic and Medicinal Chemistry**, v. 29, p. 115827, 2021.

CLAYTON, C. Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. **Open Biology**, v. 9, n. 6, 2019.

COULOMBE, R. et al. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. **EMBO Journal**, v. 15, n. 20, p. 5492–5503, 1996.

COURA, J. R. The main sceneries of chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 277–282, 2015.

DANA, D.; PATHAK, S. K. **A review of small molecule inhibitors and functional probes of human cathepsin L**. [s.l: s.n.]. v. 25

DE MAGALHÃES, C. S.; BARBOSA, H. J. C.; DARDENNE, L. E. A genetic algorithm for the ligand-protein docking problem. **Genetics and Molecular Biology**, v. 27, n. 4, p. 605–610, 2004a.

DE MAGALHÃES, C. S.; BARBOSA, H. J. C.; DARDENNE, L. E. Selection-insertion schemes in genetic algorithms for the flexible ligand docking problem. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 3102, n. ii, p. 368–379, 2004b.

DE SOUZA, D. DO S. M. et al. Anatomopathological aspects of acute chagas myocarditis by oral

transmission. **Arquivos Brasileiros de Cardiologia**, v. 107, n. 1, p. 77–80, 2016.

DEMIDYUK, I. V. et al. Propeptides as modulators of functional activity of proteases. **Biomolecular Concepts**, v. 1, n. 3–4, p. 305–322, 2010.

DOCAMPO, N. L. Z.-H. L. S. N. R. CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in *Trypanosoma cruzi* Reveals Their Role in Flagellar Attachment. **mBio**, v. 6, n. 4, p. 1012–1015, 2015.

DUMONTEIL, E.; HERRERA, C. Ten years of Chagas disease research : Looking back to achievements , looking ahead to challenges. p. 1–5, 2017.

EDER, J.; FERSHT, A. R. Pro-sequence-assisted protein folding. **Molecular Microbiology**, v. 16, n. 4, p. 609–614, 1995.

EL-SAYED, ET. AL. The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease. **Science**, v. 309, n. 5733, p. 399–400, 2005.

FERREIRA, G. A. P. L. H. S. S. C. W. L. C. M. F. S. V. W. L. M. A. D. L. C. D. A. D. A. M. A. C. R. A. N. C. R. C. K. R. L. E. R. C. AND. Benzimidazole inhibitors of the major cysteine protease of *Trypanosoma brucei*. **Future Medicinal Chemistry**, v. 11, n. 13, p. 1537–1551, 2019.

FILLBRUNN, A. et al. KNIME for reproducible cross-domain analysis of life science data. **Journal of Biotechnology**, v. 261, p. 149–156, 2017.

FRIESNER, R. A. et al. Extra precision glide: *Docking* and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. **Journal of Medicinal Chemistry**, v. 49, n. 21, p. 6177–6196, 2006.

GARCIA, M. P. et al. Characterisation of a *Trypanosoma cruzi* acidic 30 kDa cysteine protease. **Molecular and Biochemical Parasitology**, v. 91, n. 2, p. 263–272, 1998.

GAULTON, A. et al. ChEMBL: A large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, v. 40, n. D1, p. 1100–1107, 2012.

GERBABA, T. K.; GEDAMU, L. Cathepsin B gene disruption induced leishmania donovani proteome remodeling implies cathepsin B role in secretome regulation. **PLoS ONE**, v. 8, n. 11, 2013.

GILBERT, M. C. F. D. H. A. H. F. M. A. J. F. D. W. G. K. D. R. M. D. R. L. S. T. P. G. W. S. W. AND I. H. Anti trypanosomatid drug discovery: an ongoing challenge and a continuing need. **Nature Microbiology**, v. 15, p. 14, 2017.

GLOVER, J. J. AND D. H. Microhomology-mediated deletion and gene conversion in African trypanosomes. **Nucleic Acids Research**, v. 39, n. 4, p. 1372–1380, 2011.

GONZALEZ-LEAL, I. J. et al. Cathepsin B in Antigen-Presenting Cells Controls Mediators of the Th1 Immune Response during Leishmania major Infection. **PLoS Neglected Tropical Diseases**, v. 8, n. 9, 2014.

GREENSPAN, P. D. et al. Identification of dipeptidyl nitriles as potent and selective inhibitors of cathepsin B through structure-based drug design. **Journal of Medicinal Chemistry**, v. 44, n. 26, p. 4524–4534, 2001.

GREENWOOD, J. R. et al. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. **Journal of Computer-Aided Molecular Design**, v. 24, n. 6–7, p. 591–604, 2010.

- GUPTA, D. et al. CRISPR-Cas9 system: A new-fangled dawn in gene editing. **Life Sciences**, v. 232, n. April, p. 116636, 2019.
- GUPTA, R.; BRUNAK, S. Prediction of glycosylation across the human proteome and the correlation to protein function. **Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing**, v. 322, p. 310–322, 2002.
- HART, T. N.; READ, R. J. A multiple-start Monte Carlo *docking* method. **Proteins: Structure, Function, and Bioinformatics**, v. 13, n. 3, p. 206–222, 1992.
- HAYES, J. M.; STEIN, M.; WEISER, J. Accurate Calculations of Ligand Binding Free Energies: Chiral Separation with Enantioselective Receptors. **Journal of Physical Chemistry A**, v. 108, n. 16, p. 3572–3580, 2004.
- HERREROS-CABELLO, A. et al. Trypanosoma cruzi genome: Organization, multi-gene families, transcription, and biological implications. **Genes**, v. 11, n. 10, p. 1–26, 2020.
- HIGUERA, S. L.; GUHL, F.; RAMÍREZ, J. D. Identification of Trypanosoma cruzi Discrete Typing Units (DTUs) through the implementation of a High-Resolution Melting (HRM) genotyping assay. **Parasites and Vectors**, v. 6, n. 1, p. 1–6, 2013.
- HOCHSTRASSER, M. R. W. E. G. A. B. J.-C. S. K. L. W. R. D. A. AND D. F. Protein Identification and Analysis Tools in the ExPASy Server. **Methods in Molecular Biology**, v. 23, n. 5, p. 571- 607., 2005.
- HOOK, V. et al. **Cathepsin B in neurodegeneration of Alzheimer’s disease, traumatic brain injury, and related brain disorders** *Biochimica et biophysica acta. Proteins and proteomics*, 2020.
- IDEKER, P. S. A. M. O. O. N. S. B. J. T. W. D. R. N. A. B. S. AND T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 22, p. 426, 2003.
- INVITROGEN, THERMO FISHER SCIENTIFIC. BL21-AI One Shot® Chemically Competent E. coli. **Methods in Enzymology**, v. 154, n. 25, p. 2006, 2006.
- IRWIN, J. J. et al. Zinc12. **Chemical Information and Modeling**, v. 52, n. 7, p. 1757–1768, 2012.
- IRWIN, J. J.; SHOICHET, B. K. for Virtual Screening. **Journal of chemical information and modeling**, v. 45, n. 1, p. 177–182, 2005.
- JACKSON, A. S.-A. A. A. N. S.-D. B. T.-M. J. G. I P. Y. Control and management of congenital Chagas disease in Europe and other non-endemic countries : current policies and practices. **Tropical Medicine and International Health**, v. 00, n. 00, p. 1–7, 2016.
- JIA, C. et al. **Endothelial cell pyroptosis plays an important role in Kawasaki disease via HMGB1/RAGE/cathepsin B signaling pathway and NLRP3 inflammasome activation** *Cell Death and Disease*, 2019.
- JÍLKOVÁ, A. et al. Druggable Hot Spots in the Schistosomiasis Cathepsin B1 Target Identified by Functional and Binding Mode Analysis of Potent Vinyl Sulfone Inhibitors. **ACS Infectious Diseases**, v. 7, n. 5, p. 1077–1088, 2021.
- JOSEPH, B. C.; PICHAIMUTHU, S.; SRIMEENAKSHI, S. An Overview of the Parameters for Recombinant Protein Expression in Escherichia coli. **Journal of Cell Science & Therapy**, v. 06, n. 05, 2015.
- JR, A. R.; MARIN-NETO, J. A.; RASSI, A. Chronic Chagas cardiomyopathy : a review of the main pathogenic mechanisms and the efficacy of aetiological treatment following the BENznidazole Evaluation for Interrupting Trypanosomiasis ( BENEFIT ) trial. v. 112, n. March, p. 224–235, 2017.

- JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, 2021.
- JUNQUEIRA, A. C. V.; DEGRAVE, W.; BRANDÃO, A. Minicircle organization and diversity in *Trypanosoma cruzi* populations. **Trends in Parasitology**, v. 21, n. 6, p. 270–272, 2005.
- KANE, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. **Current Opinion in Biotechnology**, v. 6, n. 5, p. 494–500, 1995.
- KAPPELHOFF, R. et al. **Overview of transcriptomic analysis of all human proteases, non-proteolytic homologs and inhibitors: Organ, tissue and ovarian cancer cell line expression profiling of the human protease degradome by the CLIP-CHIP™ DNA microarray.** [s.l.] Elsevier B.V., 2017. v. 1864
- KATSUNO, K. et al. Hit and lead criteria in drug discovery for infectious diseases of the developing world. **Nature Reviews Drug Discovery**, v. 14, n. 11, p. 751–758, 2015.
- KAZLAUSKAS, E. I. A. V. AND A. Cathepsin B Regulates the Intrinsic Angiogenic Threshold of Endothelial Cells. **Molecular Biology of the Cell**, v. 16, p. 3488–3500, 2005.
- KERR, I. D. et al. Crystal structures of TbCatB and rhodesain, potential chemotherapeutic targets and major cysteine proteases of *Trypanosoma brucei*. **PLoS Neglected Tropical Diseases**, v. 4, n. 6, p. 1–10, 2010.
- KITCHEN, D. B. et al. *Docking* and scoring in virtual screening for drug discovery: Methods and applications. **Nature Reviews Drug Discovery**, v. 3, n. 11, p. 935–949, 2004.
- KLAUSEN, M. S. et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. **bioRxiv**, n. 1, 2018.
- KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. W1, p. 294–297, 2012.
- LAEMMLI. © 1970 Nature Publishing Group. **Nature Publishing Group**, v. 228, p. 726–734, 1970.
- LANDER, N. et al. CRISPR/Cas9-mediated endogenous C-terminal tagging of *Trypanosoma cruzi* genes reveals the acidocalcisome localization of the inositol 1,4,5-trisphosphate receptor. **Journal of Biological Chemistry**, v. 291, n. 49, p. 25505–25515, 2016.
- LANDER, N.; CHIURILLO, M. A.; DOCAMPO, R. Genome Editing by CRISPR/Cas9: A Game Change in the Genetic Manipulation of Protists. **The Journal of eukaryotic microbiology**, v. 63, n. 5, p. 679–690, 2016.
- LANGLEY, K. E. et al. Molecular basis of  $\beta$  galactosidase  $\alpha$  complementation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 72, n. 4, p. 1254–1257, 1975.
- LIDANI, K. C. F. et al. Chagas disease: From discovery to a worldwide health problem. **Journal of Physical Oceanography**, v. 49, n. 6, p. 1–13, 2019.
- MACH, L. et al. Proteolytic processing and glycosylation of cathepsin B. The role of the primary structure of the latent precursor and of the carbohydrate moiety for cell-type-specific molecular forms of the enzyme. **Biochemical Journal**, v. 282, n. 2, p. 577–582, 1992.
- MACIEL, A. S. A. P. F. J. A. M. N. B. C. Epidemiologia no século XXI e aspectos clínicos da doença de Chagas crônica. **Revista da SOCESP - Sociedade de Cardiologia do Estado de São Paulo**, v. 26, n. 4, p. 230–233, 2016.
- MACKEY, Z. B. et al. A cathepsin B-like protease is required for host protein degradation in *Trypanosoma brucei*. **Journal of Biological Chemistry**, v. 279, n. 46, p. 48426–48433, 2004.

- MADHAVI SASTRY, G. et al. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. **Journal of Computer-Aided Molecular Design**, v. 27, n. 3, p. 221–234, 2013.
- MAI, J. et al. Human procathepsin B interacts with the annexin II tetramer on the surface of tumor cells. **Journal of Biological Chemistry**, v. 275, n. 17, p. 12806–12812, 2000.
- MALIK, L. H.; SINGH, G. D.; AMSTERDAM, E. A. The epidemiology, clinical manifestations, and management of chagas heart disease. **Clinical Cardiology**, v. 38, n. 9, p. 565–569, 2015.
- MALLARI, J. P. et al. Structure-guided development of selective Tbc1D1 inhibitors. **Journal of Medicinal Chemistry**, v. 52, n. 20, p. 6489–6493, 2009.
- MANNE-GOEHLER, J. et al. Estimating the Burden of Chagas Disease in the United States. **PLoS Neglected Tropical Diseases**, v. 10, n. 11, p. 1–7, 2016.
- MCKERROW, J. H. The diverse roles of cysteine proteases in parasites and their suitability as drug targets. **PLoS Neglected Tropical Diseases**, v. 12, n. 8, p. 3–5, 2018.
- MELO-FILHO, C. C. **Identificação de novos compostos bioativos contra tripanossomatídeos a partir de estratégias integradas em Química Medicinal**. [s.l.: s.n.].
- MIJANOVIĆ, O. et al. Cathepsin B: A sword of cancer progression. **Cancer Letters**, v. 449, p. 207–214, 2019.
- MILES, J. S. I. F. AND M. Genetic Diversity and Genetic Exchange in *Trypanosoma cruzi*: Drug-resistant “Progeny” from Episomal Transformants. **Mem. Inst. Oswaldo Cruz**, v. 94, p. 189–193, 1999.
- MIRKOVIĆ, B. et al. Novel Mechanism of CathepsinB Inhibition by Antibiotic Nitroxoline and Related Compounds. **ChemMedChem**, v. 6, n. 8, p. 1351–1356, 2011.
- MIZUOCHI, T. et al. Both cathepsin B and cathepsin D are necessary for processing of ovalbumin as well as for degradation of class II MHC invariant chain. **Immunology Letters**, v. 43, n. 3, p. 189–192, 1994.
- MOHAMED, M. M.; SLOANE, B. F. Cysteine cathepsins: Multifunctional enzymes in cancer. **Nature Reviews Cancer**, v. 6, n. 10, p. 764–775, 2006.
- MOITESSIER, N. et al. Towards the development of universal, fast and highly accurate *docking/scoring* methods: A long way to go. **British Journal of Pharmacology**, v. 153, n. SUPPL. 1, p. 7–26, 2008.
- MÖLLER, S.; CRONING, M. D. R.; APWEILER, R. Evaluation of methods for the prediction of membrane spanning regions. **Bioinformatics**, v. 17, n. 7, p. 646–653, 2001.
- MUSIL, D. et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: The structural basis for its specificity. **EMBO Journal**, v. 10, n. 9, p. 2321–2330, 1991.
- NÄGLER, D. K. et al. Major increase in endopeptidase activity of human cathepsin B upon removal of occluding *loop* contacts. **Biochemistry**, v. 36, n. 41, p. 12608–12615, 1997.
- NDAO, M. et al. Reversible cysteine protease inhibitors show promise for a chagas disease cure. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 2, p. 1167–1178, 2014.
- NEVES. **Parasitologia humana**. 11<sup>a</sup> edição ed. São Paulo: Atheneu, 2004.
- NGÔ, H. et al. Double-stranded RNA induces mRNA degradation in *trypanosoma brucei*. **Proceedings**

of the **National Academy of Sciences of the United States of America**, v. 95, n. 25, p. 14687–14692, 1998.

NÓBREGA, O. D. T. **CARACTERIZAÇÃO MOLECULAR E FUNCIONAL DA CATEPSINA TIPO B DE TRYPANOSOMA CRUZI**. [s.l.] Universidade de Brasília, 2001.

NOVY, R. et al. Overcoming the codon bias of *E. coli* for enhanced protein expression. **inNovations**, n. 12, p. 4–6, 2001.

NOZAKI, T.; CROSS, G. A. M. Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 75, n. 1, p. 55–67, 1995.

O'BRIEN, T. C. et al. **A parasite cysteine protease is key to host protein degradation and iron acquisition** **Journal of Biological Chemistry**, 2008.

PATEL, S. et al. Cathepsins: Proteases that are vital for survival but can also be fatal. **Biomedicine and Pharmacotherapy**, v. 105, n. April, p. 526–532, 2018.

PENG, D. et al. CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. **mBio**, v. 6, n. 1, p. 1–11, 2015.

PEREIRA, G. A. N. et al. Benzimidazole inhibitors of the major cysteine protease of *Trypanosoma brucei*. **Future Medicinal Chemistry**, v. 11, n. 13, p. 1537–1551, 2019.

POLANSKI, J.; TKOCZ, A.; KUCIA, U. Beware of ligand efficiency (LE): Understanding LE data in modeling structure-activity and structure-economy relationships. **Journal of Cheminformatics**, v. 9, n. 1, p. 1–8, 2017.

PUNGERČAR, J. R. et al. Autocatalytic processing of procathepsin B is triggered by proenzyme activity. **FEBS Journal**, v. 276, n. 3, p. 660–668, 2009.

RAKSAKULTHAI, R.; HAARD, N. F. Exopeptidases and Their Application to Reduce Bitterness in Food: A Review. **Critical Reviews in Food Science and Nutrition**, v. 43, n. 4, p. 401–445, 2003.

RAO, C. S. G. AND J. S. Cathepsin B as a Cancer Target. **Expert Opin Ther Targets.**, v. 17, n. 3, p. 281–291, 2013.

RAO, J. S. Molecular mechanisms of glioma invasiveness: The role of proteases. **Nature Reviews Cancer**, v. 3, n. 7, p. 489–501, 2003.

RASID, O. et al. Cathepsin B-Deficient Mice Resolve *Leishmania major* Inflammation Faster in a T Cell-Dependent Manner. **PLoS Neglected Tropical Diseases**, v. 10, n. 5, p. 1–22, 2016.

REDECKE, L. et al. Natively Inhibited *Trypanosoma brucei* Cathepsin B Structure Determined by Using an X-ray Laser. v. 339, n. September 2012, p. 227–230, 2013.

RESTREPO, G. C. **CARACTERIZAÇÃO BIOQUÍMICA DA CATEPSINA B DE TRYPANOSOMA CRUZI**. [s.l.] Universidade de Brasília, 2001.

RESTREPO, G. E. C. **Cinética de atividade e expressão da catepsina B e caracterização bioquímica de uma leucil aminopeptidase de *Trypanosoma cruzi***. [s.l.] Universidade de Brasília, 2005.

RICO, E. et al. Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. **Scientific Reports**, v. 8, n. 1, p. 1–10, 2018.

ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server.

**Nucleic Acids Research**, v. 42, n. W1, p. 320–324, 2014.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: A unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725–738, 2010.

SAJID, M.; MCKERROW, J. H. Cysteine proteases of parasitic organisms. **Molecular and Biochemical Parasitology**, v. 120, n. 1, p. 1–21, 2002.

SANDER, T. et al. DataWarrior: An open-source program for chemistry aware data visualization and analysis. **Journal of Chemical Information and Modeling**, v. 55, n. 2, p. 460–473, 2015.

SANTOS, K. B. et al. Highly Flexible Ligand *Docking*: Benchmarking of the DockThor Program on the LEADS-PEP Protein-Peptide Data Set. **Journal of Chemical Information and Modeling**, v. 60, n. 2, p. 667–683, 2020.

SARTORI, G. R. et al. **Ligand-induced conformational selection predicts the selectivity of cysteine protease inhibitors** PLoS ONE, 2019.

SCHMITZ, J. et al. Cathepsin B Inhibitors: Combining Dipeptide Nitriles with an Occluding *Loop* Recognition Element by Click Chemistry. **ACS Medicinal Chemistry Letters**, v. 7, n. 3, p. 211–216, 2016.

SCHMITZ, J. et al. Cathepsin B: Active site mapping with peptidic substrates and inhibitors. **Bioorganic and Medicinal Chemistry**, v. 27, n. 1, p. 1–15, 2019.

SCHÖNING-STIERAND, K. et al. ProteinsPlus: Interactive analysis of protein–ligand binding interfaces. **Nucleic Acids Research**, v. 48, n. W1, p. W48–W53, 2020.

SCHRÖDINGER. **Protein Preparation Wizard**. Schrödinge ed. New York: [s.n.].

SCHRÖDINGER RELEASE, L. **No Title**, 2021.

SCHWEDE, T. et al. SWISS-MODEL: An automated protein homology-modeling server. **Nucleic Acids Research**, v. 31, n. 13, p. 3381–3385, 2003.

SELZER, P. M. et al. *Leishmania major*: Molecular modeling of cysteine protease and prediction of new nonpeptide inhibitors. **Experimental Parasitology**, v. 87, n. 3, p. 212–221, 1997.

SHIKANAI-YASUDA, M. A.; CARVALHO, N. B. Oral Transmission of Chagas Disease. v. 54, p. 845–852, 2012.

SILVA, F. S. A importância hematofágica e parasitológica da saliva dos insetos hematófagos. **Revista Tropica: Ciências Agrárias e Biológicas**, p. 3–17, 2009.

SIQUEIRA-NETO, J. L. et al. Cysteine proteases in protozoan parasites. **PLoS Neglected Tropical Diseases**, v. 12, n. 8, p. 1–20, 2018.

SLOANE, N. A. AND B. F. Cathepsin B: Multiple roles in cancer. **Proteomics Clin Appl.**, v. 8, n. 0, p. 427–437, 2014.

SOLÍS-OVIEDO, R. L. et al. Biochemical, Cellular, and Immunologic Aspects during Early Interaction between *Trypanosoma cruzi* and Host Cell. **Chagas Disease - Basic Investigations and Challenges**, 2018.

STANAWAY, J. D.; ROTH, G. The Burden of Chagas Disease. **Global Heart**, v. 10, n. 3, p. 139–144, 2015.

STEENTOFT, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell

- technology. **EMBO Journal**, v. 32, n. 10, p. 1478–1488, 2013.
- STERLING, T.; IRWIN, J. J. ZINC 15 - Ligand Discovery for Everyone. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2324–2337, 2015.
- STIERAND, K.; MAASS, P. C.; RAREY, M. Molecular complexes at a glance: Automated generation of two-dimensional complex diagrams. **Bioinformatics**, v. 22, n. 14, p. 1710–1716, 2006.
- STUDER, G. et al. QMEANDisCo—distance constraints applied on model quality estimation. **Bioinformatics**, v. 36, n. 6, p. 1765–1771, 2020.
- SUN, Y. et al. Translational Study of Alzheimer’s Disease (AD) biomarkers from brain tissues in A $\beta$ PP/PS1 mice and serum of AD patients. **Journal of Alzheimer’s Disease**, v. 45, n. 1, p. 269–282, 2015.
- SZKLARCZYK, D. et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. **Nucleic Acids Research**, v. 47, n. D1, p. D607–D613, 2019.
- TEIXEIRA, A. R. L.; NASCIMENTO, R. J.; STURM, N. R. Teixeira\_et al\_Evolution and pathology in Chagas disease - A Review.pdf. v. 101, n. August, p. 463–491, 2006.
- TIBSHIRANI, B. E. AND R. J. **An Introduction to the Bootstrap**. [s.l.] Springer-Science+Business Media, B.V., 1993.
- TOMASINI, N.; DIOSQUE, P. Evolution of Trypanosoma cruzi: Clarifying hybridisations, mitochondrial introgressions and phylogenetic relationships between major lineages. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 403–413, 2015.
- TRIFINOPOULOS, J. et al. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic acids research**, v. 44, n. W1, p. W232–W235, 2016.
- TROTT, O.; OLSON, A. J. AutoDock Vina: Improving the speed and accuracy of *docking* with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. NA-NA, 2010.
- TURK, D. et al. Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. **FEBS Letters**, v. 384, n. 3, p. 211–214, 1996.
- TURK, V. et al. Cysteine cathepsins: From structure, function and regulation to new frontiers. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1824, n. 1, p. 68–88, 2012.
- ULLMANN, A.; JACOB, F.; MONOD, J. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the  $\beta$ -galactosidase structural gene of Escherichia coli. **Journal of Molecular Biology**, v. 24, n. 2, p. 339–343, 1967.
- ULLU, E. et al. RNA interference: Advances and questions. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 357, n. 1417, p. 65–70, 2002.
- VAN DEN HAZEL, H. B.; KIELLAND-BRANDT, M. C.; WINTHER, J. R. The propeptide is required for in vivo formation of stable active yeast proteinase A and can function even when not covalently linked to the mature region. **Journal of Biological Chemistry**, v. 268, n. 24, p. 18002–18007, 1993.
- VAZQUEZ, M. P. The genetics and genomics of Trypanosoma cruzi. 2007.
- VAZQUEZ, M. P.; LEVIN, M. J. Functional analysis of the intergenic regions of TcP2 $\beta$  gene loci allowed

the construction of an improved *Trypanosoma cruzi* expression vector. **Gene**, v. 239, n. 2, p. 217–225, 1999.

VERHAMME, I. M.; LEONARD, S. E.; PERKINS, R. C. **Proteases: Pivot points in functional proteomics**. [s.l.: s.n.]. v. 1871

VERLI, H. **Bioinformática: da biologia à flexibilidade molecular**. 1ª edição ed. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2014.

VIEIRA, C. B.; PRAÇA, Y. R.; LUÍS, K. Triatomines : Trypanosomatids , Bacteria , and Viruses Potential Vectors ? v. 8, n. November, p. 1–12, 2018.

VIZOVIŠEK, M.; FONOVIĆ, M.; TURK, B. Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond. **Matrix Biology**, v. 75–76, p. 141–159, 2019.

WARDAH, W. et al. Protein secondary structure prediction using neural networks and deep learning: A review. **Computational Biology and Chemistry**, v. 81, n. July, p. 1–8, 2019.

WASZKOWYCZ, B.; CLARK, D. E.; GANCIA, E. Outstanding challenges in protein-ligand *docking* and structure-based virtual screening. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 2, p. 229–259, 2011.

WATERHOUSE, A. et al. SWISS-MODEL: Homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. W1, p. W296–W303, 2018.

WEI, B. et al. Discovery of Peptidomimetic Antibody-Drug Conjugate Linkers with Enhanced Protease Specificity. **Journal of Medicinal Chemistry**, v. 61, n. 3, p. 989–1000, 2018.

WHO. **Control of Neglected Tropical Diseases (NTD) World Health Organization**. Disponível em: <[https://www.who.int/docs/default-source/ntds/chagas-disease/chagas-2018-cases.pdf?sfvrsn=f4e94b3b\\_2](https://www.who.int/docs/default-source/ntds/chagas-disease/chagas-2018-cases.pdf?sfvrsn=f4e94b3b_2)>.

WHO. **Chagas disease (American Trypanosomiasis)**. Disponível em: <[https://www.who.int/en/news-room/fact-sheets/detail/chagas-disease-\(americantrypanosomiasis\)](https://www.who.int/en/news-room/fact-sheets/detail/chagas-disease-(americantrypanosomiasis))>.

WILKINSON, S. R. et al. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. 2008.

WILLIAMS, C. J. et al. *Molprobit*: More and better reference data for improved all-atom structure validation. **Protein Science**, v. 27, n. 1, p. 293–315, 2018.

XIONG, S. et al. Solubility of disulfide-bonded proteins in the cytoplasm of *Escherichia coli* and its “oxidizing” mutant. **World Journal of Gastroenterology**, v. 11, n. 7, p. 1077–1082, 2005.

XU, D. et al. Evaluation of high efficiency gene knockout strategies for *Trypanosoma cruzi*. **BMC Microbiology**, v. 9, p. 1–10, 2009.

YADATI, T. et al. The Ins and Outs of Cathepsins: Physiological Function and Role in Disease Management. **Cells**, v. 9, n. 7, 2020.

YAGOUBAT, A. et al. Gene Editing in Trypanosomatids: Tips and Tricks in the CRISPR-Cas9 Era. **Trends in Parasitology**, v. 36, n. 9, p. 745–760, 2020.

YANG, J. et al. The I-TASSER suite: Protein structure and function prediction. **Nature Methods**, v. 12, n. 1, p. 7–8, 2014.

- YANG, J. et al. Improved protein structure prediction using predicted interresidue orientations. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, n. 3, p. 1496–1503, 2020.
- YANG, J.; ZHANG, Y. I-TASSER server: New development for protein structure and function predictions. **Nucleic Acids Research**, v. 43, n. W1, p. W174–W181, 2015.
- ZDOBNOV, E. M. et al. OrthoDB in 2020: Evolutionary and functional annotations of orthologs. **Nucleic Acids Research**, v. 49, n. D1, p. D389–D393, 2021.
- ZHANG, C.; QUAN, R.; WANG, J. Development and application of CRISPR/Cas9 technologies in genomic editing. **Human molecular genetics**, v. 27, n. R2, p. R79–R88, 2018.
- ZHANG, H. et al. Advances in understanding the expression and function of dipeptidyl peptidase 8 and 9. **Molecular Cancer Research**, v. 11, n. 12, p. 1487–1496, 2013.
- ZHANG, H. D. Y. J. AND Y. Protein Structure Prediction. **Int J Mod Phys B.**, v. 32, n. 18, 2018.
- ZHANG, J. et al. Quantitative proteomics of cerebrospinal fluid from patients with Alzheimer disease. **Journal of Alzheimer's Disease**, v. 7, n. 2, p. 125–133, 2005.
- ZHAO, M.-M. et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development 2 3. **medRxiv**, p. 2020.10.25.20218990, 2020.
- ZHENG, W. et al. I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. **Future Generation Computer Systems**, v. 99, p. 73–85, 2019.