



**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CEILÂNDIA
CURSO DE FARMÁCIA**

MATHEUS DA SILVA ROSA

**Modelagem Molecular e garimpagem de inibidores específicos da
Oligopeptidase B de *Trypanosoma cruzi***

BRASÍLIA, 2021.

MATHEUS DA SILVA ROSA

**Modelagem Molecular e garimpagem de inibidores específicos da
Oligopeptidase B de *Trypanosoma cruzi*.**

Monografia de Conclusão de Curso
apresentada como requisito parcial para
obtenção do grau de Farmacêutico, na
Universidade de Brasília, Faculdade de
Ceilândia.

Orientadora: Profa Dra. Flávia Nader Motta

Co-orientadora: Profa Dra. Melina Mottin

Profa Dra. Camila Alves Arede

BRASÍLIA, 2021.

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

dR788m da Silva Rosa, Matheus
Modelagem Molecular e garimpagem de inibidores
específicos da Oligopeptidase B de Trypanosoma cruzi /
Matheus da Silva Rosa; orientador Flávia Nader Motta; co
orientador Camila Alves Areda. -- Brasília, 2021.
36 p.

Monografia (Graduação - Farmácia) -- Universidade de
Brasília, 2021.

1. Doença de Chagas. 2. protease. 3. Oligopeptidase B. 4.
docking molecular. 5. inibidores. I. Nader Motta, Flávia,
orient. II. Alves Areda, Camila, co-orient. III. Título.

MATHEUS DA SILVA ROSA

**Modelagem Molecular e garimpagem de inibidores específicos da
Oligopeptidase B de *Trypanosoma cruzi*.**

BANCA EXAMINADORA

Orientadora: Profa. Dra. Flávia Nader Motta CBBS-FCE

Co-Orientadora: Profa. Dra. Camila Alves Arede Farmácia-FCE

Prof. Dr. Luis Alexandre Muehlmann CBBS-FCE

Profa. Dra. Carla Nunes de Araújo CBBS-FCE

BRASÍLIA, 2021.

Agradecimentos

Gostaria de agradecer primeiramente a Deus pelo fôlego de vida e a oportunidade de vivenciar todos os momentos na Universidade de Brasília. Agradeço a minha digníssima mãe Rozilene Rosa, que desde sempre me ensinou e apoiou. Muito obrigado à minha orientadora Dra. Flávia Motta pela paciência, pelos ensinamentos e dedicação a me orientar neste trabalho. Agradeço também a minha co-orientadora Dra. Camila Areda, a pós-doc do Laboratório de interação patógeno-hospedeiro, Dra Melina Mottin, e Patrício Stegmann por me auxiliarem de diversas formas no decorrer deste período. Pelos cinco anos vividos nesta Universidade, agradeço a Universidade de Brasília pelo conhecimento, oportunidades e vivência com pessoas incríveis. Obrigado ao Prof. Dr. Bruno Junior Neves do LabMol juntamente a Universidade Federal de Goiás por me auxiliarem com ferramentas para realização deste trabalho.

Lista de abreviações

DTN - Doenças tropicais negligenciadas;

IP3 - Inositol trifosfato;

LE - *ligand efficiency*

OPBLm - Oligopeptidase B de *Leishmania major*;

OPBTb - Oligopeptidase B de *Trypanosoma brucei*;

OPBTc - Oligopeptidase B de *Trypanosoma cruzi*;

PDB - *Protein Data Bank*;

POP - Prolil Oligopeptidase;

pLDDT - *Per-residue confidence score*;

QSAR - Relação estrutura-atividade quantitativa;

Lista de figuras

Figura 1 - (A) Sobreposição estrutural da OPBLm e da OPBTb. (B) Os bolsos de ligação S1, S2 e S3 das OPB

Figura 2 - Representação esquemática do papel do OPB na doença de Chagas.

Figura 3 - Diagrama dos procedimentos e triagem.

Figura 4 - Estrutura tridimensional da OPBTc gerada pelo servidor AlphaFold (Q4CW30) com sítio de ligação do inibidor antipaína.

Figura 5 - (A) Validação do procedimento de *redocking* para o sítio de ligação da antipaína na OPBTb. (B) Ligações intermoleculares do *redocking* da antipaína com a OPBTb.

Figura 6 - (A) *Docking* da antipaína com a OPBTc. (B) Ligações intermoleculares do *docking* da antipaína com a OPBTc.

Figura 7 - Diagrama 2D de interações intermoleculares dos compostos 1, 2, 3 e 4 com a OPBTc.

Figura 8 - Diagrama 2D de interações intermoleculares dos compostos 6 e 17 com a OPBTc.

Figura 9 - Diagrama 2D de interações intermoleculares dos compostos 19, 22, 27 e 34 com a OPBTc.

Resumo

A oligopeptidase B de *Trypanosoma cruzi* (OPBTc) participa no processo de infecção do parasito e tem sido cada vez mais mencionada como um importante fator de virulência em tripanossomatídeos, uma vez que não possui ortólogos em mamíferos. A identificação de substâncias com efeitos sobre a atividade enzimática da OPBTc pode servir como molde para o desenho racional de inibidores eficazes e seguros, e portanto, com maior possibilidade de sucesso terapêutico. Neste trabalho, realizamos diferentes abordagens computacionais para obtenção da estrutura tridimensional da enzima, *docking* molecular e *virtual screening* para identificação de possíveis inibidores para OPBTc. O modelo tridimensional da OPBTc proveniente do servidor AlphaFold foi utilizado para busca de possíveis inibidores utilizando a biblioteca de compostos *Serine Proteases Inhibitors Library* (ChemDiv). Para tal, um estudo de *docking* molecular no sítio ativo da enzima foi realizado. Dos quase 33.000 compostos iniciais, 6.004 foram priorizados. Posteriormente, foi feita uma análise de eficiência de ligantes e, nesta etapa, 2.504 compostos foram priorizados. Realizamos a triagem virtual desses compostos através de um modelo de QSAR para predição de atividade contra *T. cruzi* e, neste momento, 213 compostos foram priorizados. Após a análise de grupos de similaridade estrutural, 66 compostos foram selecionados. Desses compostos, 35 passaram para a análise de inspeção visual. No entanto, os demais serão investigados futuramente. Foi possível observar que os compostos se ligam no sítio catalítico da OPBTc, próximo à tríade catalítica, e apresentam ligações de hidrogênio e interações hidrofóbicas, com resíduos semelhantes aos observados na antipaína.

Palavras-chave: Doença de Chagas, protease, Oligopeptidase B, *docking molecular*, inibidores

Abstract

Oligopeptidase B from *Trypanosoma cruzi* (OPBTc) participates in the parasite infection process and has been increasingly mentioned as an important virulence factor in trypanosomatids, since it does not have orthologs in mammals. OPBTc can serve as a template for the rational design of effective and safe inhibitors, and therefore, with a greater possibility of therapeutic success. In this work, we perform different computational approaches to obtain the three-dimensional structure of the enzyme, molecular docking and virtual screening to identify possible inhibitors for OPBTc. The three-dimensional OPBTc model generated from the AlphaFold server was used to search for possible inhibitors using Serine Proteases Inhibitors Library (ChemDiv). To this end, a study of molecular docking in the active site of the enzyme was carried out. Of the nearly 33,000 initial compounds, 6,004 were prioritized. Subsequently, an analysis of ligand efficiency was carried out and, in this step, 2,504 compounds were prioritized. We performed a virtual compound screening using a QSAR model for phenotypic data of compounds tested against *T. cruzi* and, at this time, 213 compounds were prioritized. After an analysis of structural similarity groups, 66 compounds were selected. Of these 66 compounds, 35 passed to visual inspection. Therefore we observed that those compounds interact with the enzyme catalytic site similarly to those antipain-OPBTc and OPBTb hydrogen bonds, hydrophobic and pi-stacking interactions at the same protein residues.

Keywords: Chagas disease, protease, Oligopeptidase B, molecular docking, inhibitors

SUMÁRIO

Introdução	9
Doença de Chagas	9
Oligopeptidase B (OPB)	11
Oligopeptidase B de <i>Trypanosoma cruzi</i> (OPBTc)	12
Predição de estruturas tridimensionais	15
Justificativa	16
Objetivo geral	16
Metodologia	17
Sequência primária da OPBTc	17
Predição das estruturas tridimensionais da OPBTc	17
Preparação dos compostos e proteína para cálculos de <i>docking</i>	18
<i>Docking molecular</i>	18
Filtro de eficiência de ligante	18
Filtro de QSAR fenotípico	19
Cálculo de descritores moleculares e Agrupamento de compostos	19
Análises de Inspeção Visual	20
Resultados e discussões	21
Conclusão	31
Bibliografia	32
Anexo 01 - Sequência primária	35

Introdução

Doenças atingem a sociedade de diversas maneiras e as consequências destes eventos se manifestam com pesos diferentes de acordo com o estado socioeconômico, nível de desenvolvimento do país e seu sistema de saúde. As doenças tropicais negligenciadas (DTN), como malária, doença de Chagas, febre Ebola e dengue têm maior incidência nas regiões tropicais e suas características emergentes e reemergentes vem se tornando motivo de alerta (BERMUDEZ et al., 2016; KRATZ, 2019).

Doença de Chagas

A doença de Chagas é uma infecção endêmica na maioria da América Latina, também chamada de tripanossomíase americana causada pelo protozoário *Trypanosoma cruzi*, sendo os insetos do gênero *Triatoma* o principal vetor de infecção, porém há a ocorrência de infecção pelo sangue ou material infectado com o mesmo, de forma congênita da mãe para o filho, transplante de órgãos, acidentes laboratoriais e até por consumo de alimentos não higienizados, como o açaí (HEMMIGE; TANOWITZ; SETHI, 2012). A prevalência dessa doença vem aumentando nas últimas décadas, sendo mais de 8 milhões de pessoas sofrendo com a doença e mais de 75 milhões em risco de contaminação (WHO, 2021).

O desafio do tratamento e prevenção das tripanossomíases é contínuo por não haver vacinas efetivas e os fármacos disponíveis falham ao eliminar os parasitos. Esforços para a criação de uma vacina encontram barreiras, principalmente sendo a forma com que o parasito interage com o sistema imune inato um grande desafio. Uma potencial vacina para *T. cruzi* deve incluir antígenos presentes em todas as formas do parasito e deve servir como profilaxia e ter efeito terapêutico (RIOS et al., 2019).

O ciclo de vida do *T. cruzi* é complexo, existem mais de 135 espécies de triatomíneos que são vetores potenciais do parasito. O ciclo de vida do parasito ocorre no hospedeiro vertebrado e no vetor começando com o repasto sanguíneo em um mamífero infectado ingerindo a forma tripomastigota de *T. cruzi* (JANSEN; XAVIER; ROQUE, 2018); (HEMMIGE; TANOWITZ; SETHI, 2012). No sistema digestório do inseto ocorre reprodução e desenvolvimento do parasito onde a forma tripomastigota se diferencia em forma epimastigota e se reproduz por fissão binária. Posteriormente os epimastigotas se diferenciam em tripomastigotas metacíclicos no

intestino posterior do inseto. A forma tripomastigota metacíclica é liberada nas fezes do inseto durante o repasto sanguíneo, adentrando a ferida causada durante o repasto e alcançando a corrente sanguínea. O parasito, no hospedeiro mamífero, é intracelular e após invadir a célula se diferencia em amastigota e se reproduz por fissão binária e, em sequência, há diferenciação para a forma tripomastigota ainda dentro da célula hospedeira. Eventualmente ocorre a lise da célula hospedeira e os tripomastigotas são liberados na circulação sanguínea e invadem outras células. O ciclo de infecção se completa quando as formas tripomastigotas circulantes no sangue são ingeridas pelo inseto vetor durante um outro repasto sanguíneo (HEMMIGE; TANOWITZ; SETHI, 2012).

Após a infecção há um período de incubação, seguido pela fase aguda, que pode ser assintomática, oligossintomática ou sintomática. Por aproximadamente dois meses e na maior parte dos casos os sintomas são leves e sem especificidade, logo pode ser confundido com outras doenças. Em pessoas picadas pelo inseto, os primeiros sinais são lesões na pele ou chagoma de inoculação, conhecido como sinal de Romaña. Nos casos sintomáticos também pode ocorrer febre, dores de cabeça, aumento dos gânglios linfáticos, palidez, dor muscular, abdominal ou dor no peito, assim como dificuldade de respirar (WHO, 2021).

O tratamento farmacológico acontece com dois medicamentos, benzonidazol e nifurtimox, fármacos antiparasitários que são efetivos em curar a doença em seu estágio inicial. No Brasil, somente o tratamento com benzonidazol é aceito. A eficácia de ambos os fármacos decai à medida que o tempo de infecção aumenta promovendo também maiores efeitos adversos. Pacientes adultos, especialmente sintomáticos, devem realizar o tratamento para diminuir a progressão da doença, assim como evitar a transmissão congênita em mulheres gestantes, sendo considerado tratamento precoce a gravidez. Pacientes que manifestaram novamente a infecção devido a imunossupressão ou estão no início da fase crônica também são indicados para realizar o tratamento (KRATZ, 2019).

O surgimento de novos fármacos para o tratamento de doenças negligenciadas é um desafio para a Doença de Chagas, assim como outras tripanossomíases por se conhecer pouco da farmacologia antiparasitária, quando comparado com outras doenças de impacto social ou econômico. A quimioterapia com antiparasitários é a estratégia terapêutica em progresso, assim como o guia para o desenvolvimento de novos fármacos. O primeiro passo para o

desenvolvimento de um novo fármaco é a identificação de pontos de partida químicos por meio da varredura de bibliotecas químicas usando uma abordagem baseada no alvo do fármaco ou celular. Esta abordagem celular é a mais comum com referências na literatura (BERMUDEZ et al., 2016; KRATZ, 2019).

Oligopeptidase B (OPB)

O estudo de alvos farmacológicos dentre as estruturas bioquímicas do agente infeccioso é de grande importância para o desenvolvimento de um tratamento mais específico para a doença, assim como as outras DTN. Dentro desse grupo, temos as proteases, enzimas reconhecidas pela sua participação nos processos de infecção e estabelecimento de doenças, podendo ser consideradas como potenciais fatores de virulência (BERMUDEZ et al., 2016; KRATZ, 2019; STUDER, 2020).

A oligopeptidase B (OPB), protease citosólica, pertence ao clã SC, família prolil oligopeptidase (POP), família S9 de serino-proteases, com especificidade para clivar o lado carboxi de resíduos básicos, com preferência para a arginina sobre a lisina (COETZER; GOLDRING; HUSON, 2008). Sua distribuição é restrita a bactérias Gram-negativas e positivas, eucariotos inferiores e plantas e não possui ortólogos em mamíferos (COETZER; GOLDRING; HUSON, 2008), característica importante quando se pensa nesta enzima como alvo de fármaco.

Com relação à estrutura da enzima, o primeiro modelo 3D de OPB (*Escherichia coli*) foi construído por modelagem por homologia baseada na estrutura da POP suína (GÉRCZEI; KESERÜ; NÁRAY-SZABÓ, 2000). Essa previsão foi então confirmada após a cristalização e resolução das OPB de *Leishmania major* (OPBLm) e *Trypanosoma brucei* (OPBTb) (Figura 01), mostrando uma estrutura composta por dois domínios: α/β hidrolase, onde está localizada a tríade catalítica conservada - Ser, Asp e His e o β propeller (CANNING et al., 2013; MCLUSKEY et al., 2010). Ambas OPB foram co-cristalizadas na presença do inibidor antipaína (Phe-Arg-Val-Arg) o que permitiu compreender a interação das enzimas com substratos dibásicos (CANNING et al., 2013; MCLUSKEY et al., 2010).

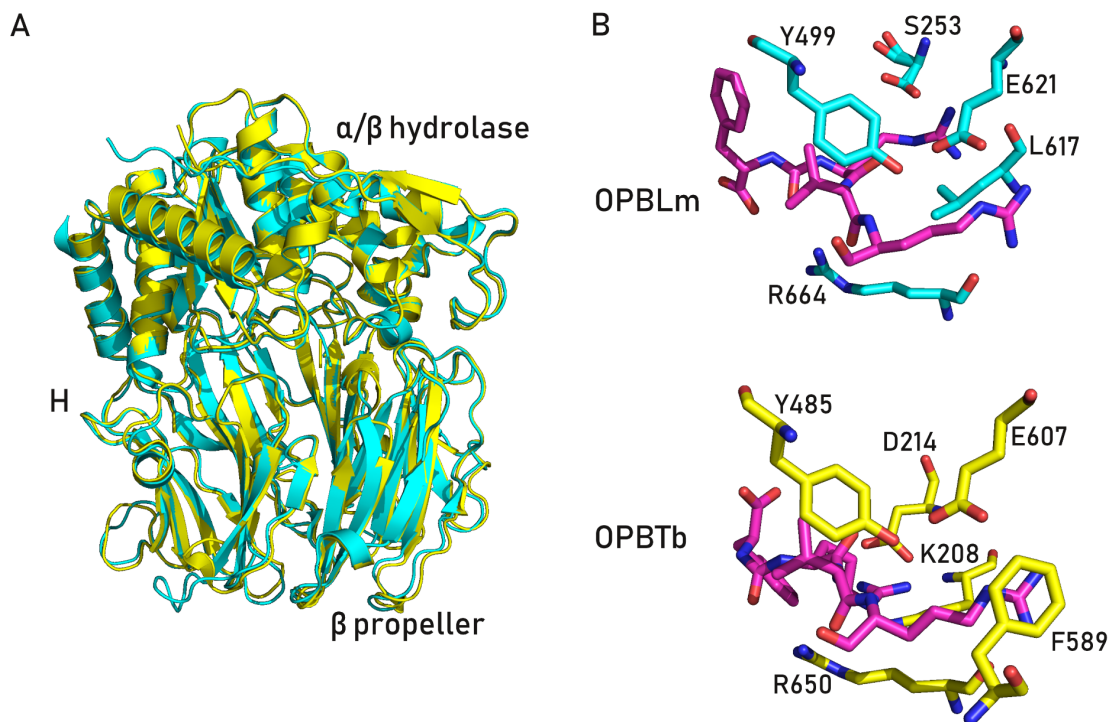


Figura 1. (A) Sobreposição estrutural da OPBLm (ciano) e da OPBTb (amarelo); H: região da dobradiça. (B) Os bolsos de ligação S1, S2 e S3 das OPB. Magenta: antipaina. OPBLm PDB: 2XE4 e OPBTb PDB: 4BP9 (Adaptada de (MORTY et al., 1999a, 1999b; MOTTA et al., 2019)).

A presença da OPB em eucariotos foi mostrada, pela primeira vez, em *T. cruzi* (SANTANA et al., 1992) e suas implicações na doença de Chagas serão descritas a seguir.

Oligopeptidase B de *Trypanosoma cruzi* (OPBTc)

A OPBTc é encontrada no citosol de tripanossomatídeos, mas seu papel fisiológico ainda não foi bem elucidado, assim como seu substrato fisiológico também não foi identificado. Estudos mostram que a oligopeptidase B está diretamente relacionada com o influxo transitório de Ca^{2+} em células de mamíferos, provocando um recrutamento e fusão de lisossomos para o local de invasão do parasito, pois, diferentemente de outros parasitos, essa organela é necessária para a entrada do *T. cruzi* nas células do hospedeiro até completar o seu ciclo de vida, que permite as condições ideais de pH, formando um vacúolo parasitóforo contendo o parasito. Sua participação na invasão celular ocorre por meio da geração de um agonista que ativa fosfolipase C da célula hospedeira resultando na formação de inositol trifosfato (IP3) e, conseqüentemente, na liberação de Ca^{2+} intracelular. No entanto, não se sabe se a produção desse agonista ocorre no citoplasma do parasito

ou no meio extracelular pela OPBTc secretada (BURLEIGH e ANDREWS, 1995) (RODRIGUEZ e col., 1995) (FERNANDES e col., 2005) (Figura 2). A OPBTc não possui por si só a capacidade de ativar esta via de sinalização em células infectadas, esta via é fundamental para infecção de células epiteliais, fibroblastos e mioblastos, mas não para macrófagos, onde a invasão celular dispensa a atividade da enzima (BURLEIGH; ANDREWS, 1995; BURLEIGH et al., 1997; ILLENBERGER et al., 2003). Segundo Caler (1998), a deleção do gene que codifica OPB resultou em uma redução significativa da capacidade de infecção pelo *T. cruzi*, e em adição, a OPBTc recombinante purificada foi capaz de restaurar a sinalização de Ca^{2+} no mutante (CALER e col., 1998).

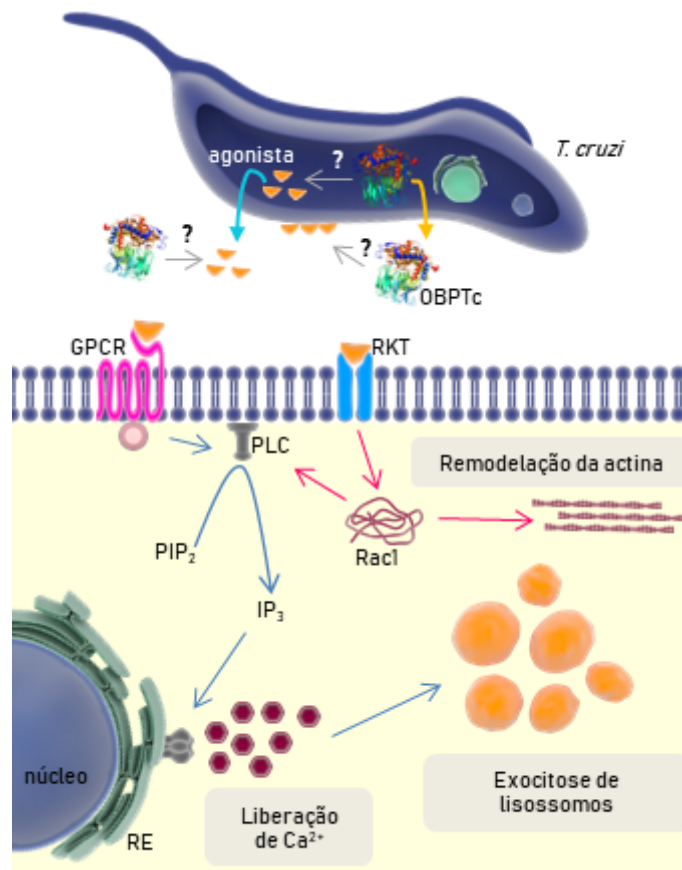


Figura 2. Representação esquemática do papel da OPB na doença de Chagas. OPBTc hidrolisa um precursor desconhecido ativando um agonista que culmina no aumento de Ca^{2+} intracelular. O agonista se liga ao receptor da célula hospedeira e ativa a fosfolipase C (PLC), levando à formação de trifosfato de inositol (IP_3) e, conseqüentemente, a mobilização de Ca^{2+} . Ainda não está claro se OPBTc interage com seu substrato dentro ou fora do parasito. ? - etapas desconhecidas; RE, retículo endoplasmático; GPCR: Receptor acoplado à proteína G; PLC, fosfolipase C; PIP_2 : fosfatidilinositol 4,5-bisfosfato; IP_3 , trifosfato de inositol; RKT, receptor do tipo tirosina quinase; Rac1: Ras-relacionada

ao substrato C3 da Toxina botulínica 1 (Adaptada de (MORTY et al., 1999a, 1999b; MOTTA et al., 2019).

A oligomerização já foi amplamente estudada para OPBTc (MOTTA et al., 2012; SANTANA et al., 1992). Foi proposta pela primeira vez porque a forma nativa da OPBTc migra, após eletroforese, como uma proteína de 80 kDa quando totalmente desnaturada e com massa molecular aparente de 120 kDa sem aquecimento prévio (MOTTA et al., 2012; SANTANA et al., 1992). Em estudos de cromatografia de exclusão por tamanho (SEC), OPBTc eluiu como um dímero de 160 kDa, mesmo quando pré-tratada em diferentes pHs, temperaturas, concentrações de sal e também na presença de DTT, sugerindo a formação de um dímero estável (MOTTA et al., 2012). Uma vez que a eletroforese e a SEC apenas estimam a massa de proteínas, Motta e colegas aplicaram a ultracentrifugação analítica (UCA), uma ferramenta precisa para a identificação de estados oligoméricos (LEBOWITZ; LEWIS; SCHUCK, 2002), e ficou, então, demonstrado, pela primeira vez, a natureza dimérica de OPBTc (MOTTA et al., 2012). O ensaio de UCA também mostrou que 90% do sistema OPBTc era formado por dímeros e o restante era provavelmente agregado, o que reforça a ideia do dímero estável (MOTTA et al., 2012). No entanto, ainda não se sabe se o monômero da enzima é ativo.

Algumas das substâncias passíveis de sofrer hidrólise pela OPB são hormônios peptídicos como fator natriurético atrial, hormônio adrenocorticotrófico, glucagon, neurotensina, angiotensina I e peptídeo vasoativo intestinal. Há também inibidores irreversíveis de serino peptidases que atuam nas OPB de tripanossomatídeos como 3,4-dicloroisocumatina, 4-(2aminoetila) benzenosulfonilfluoreto, diisopropilfluorofosfato, fenilmetanosulfonilfluoreto, leupeptina e antipaína. Os estudos relacionados à função biológica da OPB sempre foram executados com inibidores clássicos de tripsina, como a antipaína, porém não há muitos estudos buscando inibidores específicos para OPB (MORTY et al., 1999a, 1999b; MOTTA et al., 2019).

A identificação da interação entre proteína e ligante, seja um sítio ativo ou sítio inibitório é de suma importância para que seja possível entender a função de uma enzima. Sendo assim, a função biológica de uma proteína é ditada pela sua estrutura tridimensional, logo um dos motivos de se predizer a sua estrutura é obter

a informação de sua função. Porém a relação entre estrutura e função não é sempre coerente, visto que diversas famílias de proteínas possuem funções ambíguas e diferentes arranjos estruturais podem efetuar a mesma função. Devido a este fato, a análise de estruturas quaternárias e terciárias de forma separada pode explicar como estruturas em seu escopo global, estrutura quaternária, que prevalecem diferentes podem ter a mesma função quando analisadas em um escopo local, estrutura terciária (ROY; KUCUKURAL; ZHANG, 2010; WATERHOUSE et al., 2018; YANG; ROY; ZHANG, 2013; YANG et al., 2015). Neste sentido, a OPBTc é um alvo atrativo pois essa enzima não apresenta ortólogos em mamíferos e um inibidor para esta protease poderia impedir o avanço da doença sem causar efeitos adversos no paciente. Uma metodologia computacional com simulações de dinâmica molecular, acoplamento molecular, previsão de estruturas tridimensionais e estudo do processo farmacocinético são adotadas para pesquisa de possíveis novas moléculas considerando diversas bases de dados. Este processo permite diminuir custos e tempo de pesquisa (DIHARCE et al., 2019).

Predição de estruturas tridimensionais

As ferramentas de predição de estrutura 3D e função de proteínas *in silico* são diversas são usados o servidor I-TASSER, trRosetta, Alfaphold e SWISS MODEL. O servidor I-TASSER para predição de estrutura e função utiliza uma metodologia baseada na sequência terciária (enovelamento) de aminoácidos de uma proteína. O primeiro estágio de processamento é a busca por um modelo, ou seja, uma proteína que possua sequência proteica similar à outra estrutura vinculada a base de dados utilizando o método PSI-BLAST. Um perfil sequencial é criado com base em múltiplos alinhamentos da sequência homóloga para que ocorra uma comparação com a biblioteca de estruturas do Protein Data Bank (PDB) usando o método LOMETS (ROY; KUCUKURAL; ZHANG, 2010; YANG et al., 2015).

As estruturas alocadas nesta biblioteca foram analisadas pelo processo de difração de raios X, que é uma técnica não destrutiva que promove de forma detalhada as informações cristalográficas, composição química e física. Os modelos são ordenados de acordo com a similaridade e altas pontuações na comparação com a sequência de entrada. A qualidade de tal comparação é julgada por parâmetros estatísticos determinados de *C-score* que indica a proximidade entre a sequência de entrada e o modelo que foi selecionado em relação a estrutura e função (YANG; ROY; ZHANG, 2013).

O servidor SWISS MODEL utiliza uma metodologia baseada nas sequências primária e secundária de aminoácidos de uma proteína homóloga para construir o modelo. Os modelos possuem um parâmetro de Estimação de Qualidade Global do Modelo (GMQE) e Estimação de Qualidade da Estrutura Quaternária (QSQE) para que seja avaliado e selecionado o melhor modelo para a estrutura final. Um modelo tridimensional é gerado baseado no modelo e para qualificá-lo é utilizado o parâmetro estatístico QMEAN que determina a qualidade local e global da estrutura final (WATERHOUSE et al., 2018).

Os servidores trRosetta e Alphafold são livres de molde e utilizam aprendizado de máquina para fazer a predição dos modelos 3D. O trRosetta constrói modelos utilizando através de aprendizado profundo (*deep learning*), embasado em uma rede de informações, para gerar a distância entre os resíduos da proteína e também as restrições das orientações, para uma sequência de estrutura desconhecida. O servidor também pode utilizar informações de estruturas 3D homólogas para auxiliar a construção do modelo (Yang et al., 2020). O servidor Alphafold, um banco de dados de modelos 3D, desenvolvido pela DeepMind and EMBL-EBI, disponibiliza modelos de proteínas, gerados a partir de alinhamentos multi-sequenciais e deep learning (JUMPER et al., 2021).

Justificativa

Além de ser um fator de virulência do parasito, a vantagem de se utilizar a OPBTc como alvo potencial de fármacos reside na ausência de ortólogos em mamíferos. Em teoria, isso facilitaria o desenvolvimento de fármacos seletivos, isto é, com menor possibilidade de efeitos colaterais ao homem. Dentro deste panorama, o conhecimento da estrutura tridimensional da OPBTc e de seus possíveis inibidores irá acrescentar mais conhecimento sobre o seu papel na biologia do parasito e no desenvolvimento e instalação da infecção. Além disso, substâncias com comprovada atuação na OPBTc podem servir de molde para o desenho racional de inibidores mais eficazes e seguros e, portanto, com maior possibilidade de êxito terapêutico.

Objetivo geral

Obter a estrutura tridimensional da OPBTc por meio de modelagem molecular e encontrar possíveis inibidores com potencial terapêutico.

Objetivos específicos

1. Prospectar e reunir as sequências de aminoácidos relevantes para a modelagem molecular da Oligopeptidase B de *T. cruzi*.
2. Construir modelos tridimensionais da proteína utilizando os servidores SWISS MODEL, I-TASSER, trRosetta e AlphaFold.
3. Comparar modelos gerados analisando a fidelidade e otimização.
4. Executar simulações de predição da interação enzima-inibidor, através de *docking* molecular utilizando modelos tridimensionais mais otimizados.
5. Filtrar e priorizar potenciais inibidores de OPBTc, por meio da integração de diferentes abordagens computacionais.

Metodologia

O trabalho será uma pesquisa do tipo exploratória com metodologia experimental e bibliográfica, assim como técnica de coleta de dados de documentação direta e indireta. Ao final desta seção, encontra-se um fluxograma ilustrativo dos procedimentos computacionais realizados durante o trabalho (Figura 3).

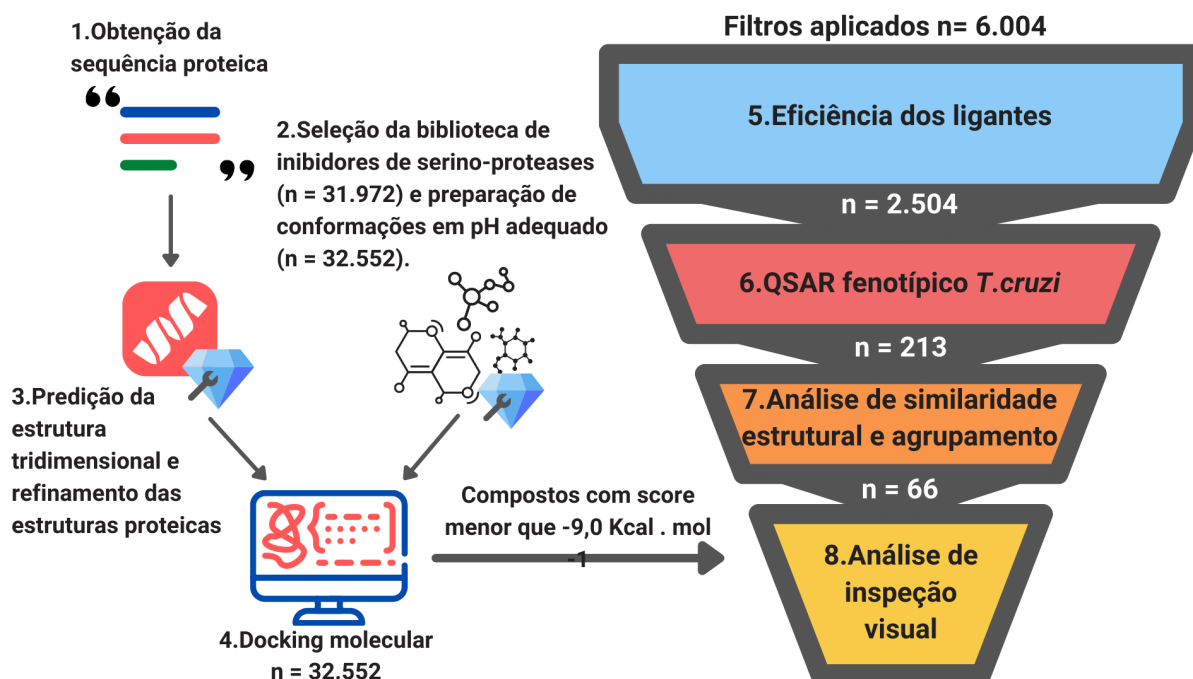


Figura 3. Fluxograma com as etapas e triagem virtual. (1) Obtenção da sequência proteica na base de dados PDB e Merops. (2) Seleção da biblioteca de inibidores de serino-protease ChemDiv e preparação das moléculas para pH $8,0 \pm 0,5$ no LigPrep. (3) Predição da estrutura tridimensional da proteína utilizando SwissModel, ITasser, trRosetta e AlphaFold, assim como refinamento das

estruturas com Galaxyrefine e preparação no Protein Wizard. (4) Submissão dos modelos e ligantes ao *Docking* molecular utilizando DockThor. (5) Triagem dos resultados utilizando filtros de eficiência dos ligantes, (6) relação estrutura-atividade quantitativa (QSAR), (7) análise de grupos de similaridade estrutural e estruturas representativas. (8) Análise de inspeção visual das estruturas representativas de melhor pontuação. Fonte: elaborado pelo autor.

Sequência primária da OPBTc

A sequência de aminoácidos da OPBTc (ID MER0002611) foi extraída do banco de dados MEROPS (RAWLINGS et al., 2018) para ser usada no SwissModel, Itasser e trRosetta. O modelo Alphafold gerado utilizou a sequência Id UnProt Q4CW30, gene Tc00.1047053503995.50.

Predição das estruturas tridimensionais da OPBTc

A partir da sequência primária, a modelagem foi realizada com a submissão do arquivo FASTA em quatro servidores: (1) SWISS-MODEL, Biozentrum, University of Basel, (2) I-TASSER, BMC Bioinformatics (YANG, 2015), (3) trRosetta, YangLab (DU, 2020) e (4) AlphaFold, Deepmind (JUMPER, 2021). Os servidores SWISS-MODEL e I-TASSER realizam a modelagem molecular baseada em molde, a partir de sequência primária e terciária, respectivamente (DENG; JIA; ZHANG, 2018). O servidor trRosetta constrói modelos utilizando deep learning, embasado em uma rede de informações, para gerar a distância entre os resíduos da proteína e também as restrições das orientações, para uma sequência de estrutura desconhecida (YANG et al., 2020). O servidor também pode utilizar informações de estruturas 3D homólogas para auxiliar a construção do modelo. Em seguida, o programa usa o minimizador Rosetta para encontrar uma conformação da cadeia principal da proteína consistente com essas restrições. O servidor Alphafold, um banco de dados de modelos 3D desenvolvido pela DeepMind and EMBL-EBI, que disponibiliza modelos 3D de proteínas, gerados a partir de alinhamentos multi-sequenciais e deep learning (JUMPER et al., 2021). Para o refinamento dos modelos foi utilizado o servidor GalaxyRefine (KO et al., 2012). O refinamento é realizado utilizando *conformational space annealing*, uma otimização das conformações de regiões locais não confiáveis (KO et al., 2012). A avaliação da qualidade dos modelos foi realizada no servidor MolProbity (DAVIS et al., 2007), que determina a qualidade de um modelo predito por meio de parâmetros que representam proteínas cristalizadas.

Preparação dos compostos e proteína para cálculos de *docking*

A biblioteca de compostos foi preparada no programa Glide, usando a ferramenta LigPrep (SCHRÖDINGER, 2015), para a correta protonação e geração de confôrmeros dos compostos em pH $8,0 \pm 0,5$.

A proteína foi preparada no programa Glide, utilizando a ferramenta *Protein Preparation Wizard* (FRIESNER et al., 2006; SASTRY et al., 2013) em pHs de $8,0 \pm 0,5$ (pH ótimo de atividade da OPBTc), 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).

***Docking* molecular**

Os cálculos de *docking* da biblioteca *Serine Proteases Inhibitors Library* (ChemDiv) contra o modelo 3D da OPBTc foram realizados utilizando o servidor DockThor (DE MAGALHÃES et al., 2014). Além disso, foi realizado o *redocking* da antipaina na estrutura da OPB de *T. brucei* (PDB ID 4BP9). As configurações do *grid* de *docking* utilizadas para OPBTc foram: coordenadas X = 0,044 Å, Y = 5,88 Å e Z = -5,37 Å com tamanho de 20 Å. O modo de *docking* utilizado foi *virtual screening*, na configuração padrão do servidor.

Filtro de eficiência de ligante

A eficiência do ligante (LE) é calculada através da razão entre a afinidade de ligação (ΔG) ou score de *docking* pelo número de átomos pesados (não hidrogênios) (n), de acordo com a Equação 1:

$$LE = \Delta G/n \quad (1)$$

Valores ideais de LE são acima de $0,3 \text{ Kcal} \cdot \text{mol}^{-1} \cdot \text{átomo não-hidrogênio}^{-1}$ (ABAD-ZAPATERO, 2007; KESERÜ; MAKARA, 2009), esses são os valores de LE aceitos para fármacos e compostos. A LE normaliza a afinidade em relação ao número de átomos, isso permite verificar a eficiência do ligante e reduzir a diferença entre compostos com estruturas pequenas em relação às estruturas grandes, que fazem mais interações nos cálculos de *docking*. Os valores de número de átomos pesados dos compostos foram calculados no programa Knime (DELEN; HELFRICH; SILIPO, 2021) e a LE foi calculada no Excel.

Filtro de QSAR fenotípico

Os modelos de relação estrutura atividade quantitativa (QSAR) fenotípicos utilizados neste trabalho foram desenvolvidos no laboratório do Prof. Dr. Bruno Junior Neves, LabMol (Universidade Federal de Goiás). Tais modelos realizam uma

predição da atividade dos compostos frente ao *T. cruzi*. Os modelos de classificação foram construídos através de aprendizado de máquina e os modelos de regressão, através de aprendizado profundo (redes neurais), utilizando dados para *T. cruzi* com concentração inibitória de 50% (IC₅₀) 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 (IC₅₀ ≤ 10 µM) e inativos (IC₅₀ > 10 µM) (KATSUNO et al., 2015). Os modelos foram empregados como filtros para predição de atividade em *T. cruzi* dos compostos da biblioteca de ligantes e o critério para filtrá-los foi a probabilidade de ser ativo igual ou superior a 0,6.

Cálculo de descritores moleculares e Agrupamento de compostos

O programa DataWarrior (SANDER et al., 2015) foi utilizado para realizar os cálculos de descritores moleculares e agrupamento dos compostos, baseado em similaridade estrutural. Um descritor molecular é utilizado para codificar as informações químicas estruturais dos compostos em valores úteis (números binários) com o SkelSpheres. Após calculados os descritores SkelSpheres para os compostos, estes foram submetidos à análise de similaridade estrutural e clusterização (agrupamento) por similaridade. Dentro de cada *cluster* (grupo) foi determinado o composto representativo, ou seja, o composto com a estrutura química principal mais simples.

Análises de Inspeção Visual

O programa Pymol (SCHRÖDINGER, 2015) foi utilizado para visualizar os resultados de *docking* e auxiliar, durante os processos anteriores, na inspeção visual, assim como para gerar as imagens das estruturas tridimensionais das proteínas com os ligantes.

As análises das interações proteína-ligante resultantes do *docking* foram processadas no servidor *ProteinPlus poseview* (STIERAND; MAASS; RAREY, 2006) para averiguação de ligação de hidrogênio, interação resultante de um átomo de hidrogênio ligado a um átomo eletronegativo e outro átomo eletronegativo, Interações hidrofóbicas, interações entre estruturas apolares que geram cargas aparentes momentâneas em seus átomos, *pi-stacking*, que é o empilhamento de moléculas apolares aromática e outras interações intermoleculares (SCHÖNING-STIERAND et al., 2020).

Resultados e discussões

Utilizando os servidores SwissModel, ITasser, trRosetta e AlphaFold, obtivemos quatro modelos da OPBTc. O modelo gerado pelo Swiss-Model utilizou como molde a estrutura 3D da Oligopeptidase B de *Trypanosoma brucei* (PDB ID: 4bp9), que apresenta identidade sequencial de 72,2% e cobertura sequencial de 84,6%. Os modelos gerados pelos servidores ITasser e trRosetta utilizaram a sequência MER0002611 na construção do modelo. O servidor Alphafold disponibiliza o modelo da OPBTc na sua base de dados e o modelo apresenta alto grau de confiabilidade estrutural. Esse modelo apresenta um *per-residue confidence score* (pLDDT) > 90 para a maioria dos resíduos. Os valores de pLDDT podem variar entre 0 a 100, sendo 100 a melhor pontuação, relacionada a resíduos bem estruturados e conformações confiáveis (Jumper et al., 2021).

A qualidade dos modelos gerados foi analisada por meio do servidor MolProbity (CHEN et al., 2010; DAVIS et al., 2007). Os parâmetros analisados levam em conta a qualidade do modelo tridimensional em características geométricas como o *clashscore*, que determina colisões entre átomos, e *Ramachandran's favored* que busca resíduos de aminoácidos em posições com maior probabilidade de estarem corretas (Tabela 1). O *MolProbity 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 de resolução de raio-X, refletindo a resolução cristalográfica de uma proteína.

Tabela 1 - Parâmetros de qualidade dos modelos gerados da OPBTc.

Modelos	ClashScore	Rama Favored	MolProbity score
Itasser	5.5	0.8778	2.62
AlphaFold	1.33	0.9691	1.04
trRosetta	5.5	0.9495	1.65
SwissModel	2.69	0.9477	1.90
Itasser refinado	8.61	0.9579	1.84
Alphafold refinado	6.31	0.9860	1.35
trRosetta refinado	8.25	0.9662	1.66

SwissModel refinado	6.8	0.9760	1.46
---------------------	-----	--------	------

Como visto na Tabela 1, o modelo tridimensional original gerado pelo servidor AlphaFold (Figura 4) apresentou a melhor pontuação Molprobity score, essa reflete um conjunto de parâmetros e, por isso, foi utilizada em todos os procedimentos computacionais. Um valor de *clashscore* próximo a zero indica que há menos colisões entre átomos na estrutura. *Ramachandran's favored* próximo de 1,0 indica a porcentagem de resíduos de aminoácido alocados em posições corretas referente a relação dos ângulos *phi* e *psi* no gráfico de Ramachandran. O modelo obtido com o servidor trRosetta também apresentou bons parâmetros de qualidade e poderá ser utilizado em trabalhos futuros. De forma geral, o refinamento das estruturas não foi eficaz para os modelos gerados pelo AlphaFold e trRosetta, apenas para SwissModel e Itasser.

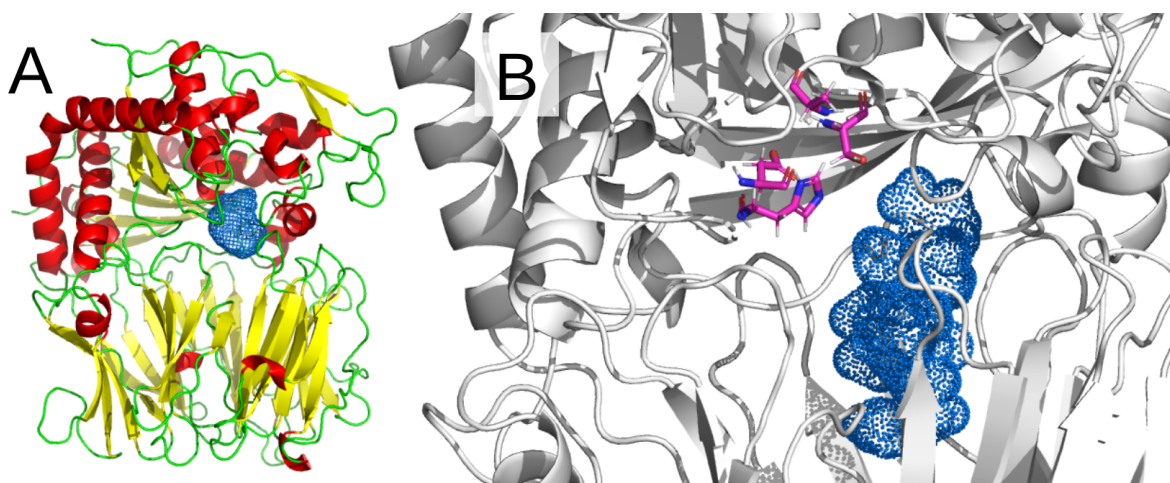


Figura 4. Estrutura tridimensional da OPBTc gerada pelo servidor AlphaFold (Q4CW30) com sítio de ligação do inibidor antipaina (azul). (A) Visão da estrutura 3D da OPBTc completa. Vermelho: Alfa-hélices; Amarelo: folhas beta; Verde: loops. (B) Zoom no sítio ativo (resíduos na representação sticks rosa) da OPBTc. Fonte: autor (DE MAGALHÃES et al., 2014; SCHÖNING-STIERAND et al., 2020; SCHRÖDINGER, 2015).

Realizamos uma validação do protocolo de *docking*, conhecida como *redocking*. Neste procedimento, a partir de uma estrutura cristalográfica da proteína com um inibidor co-cristalizado, se retira o inibidor da estrutura e realiza o cálculo de *docking* deste inibidor na proteína, para verificar se o programa acerta o modo de interação com a proteína e as interações entre eles. Dessa forma, realizamos o *redocking* da OPBTb com a antipaina (PDB ID 4bp9) e a pontuação foi de -8,06

Kcal·mol⁻¹. A análise visual indicou alta similaridade da pose de *docking* com a estrutura cristalográfica (Figura 5A). Isso mostra que o programa de *docking* DockThor está acertando o modo de interação da antipaína com a OPB *T. brucei*. Importantes ligações de hidrogênio entre a antipaína e os resíduos Arg650, Tyr482, Glu607 e Asp214 da OPBTb podem ser visualizados, assim como interações hidrofóbicas com Val651 (Figura 5B).

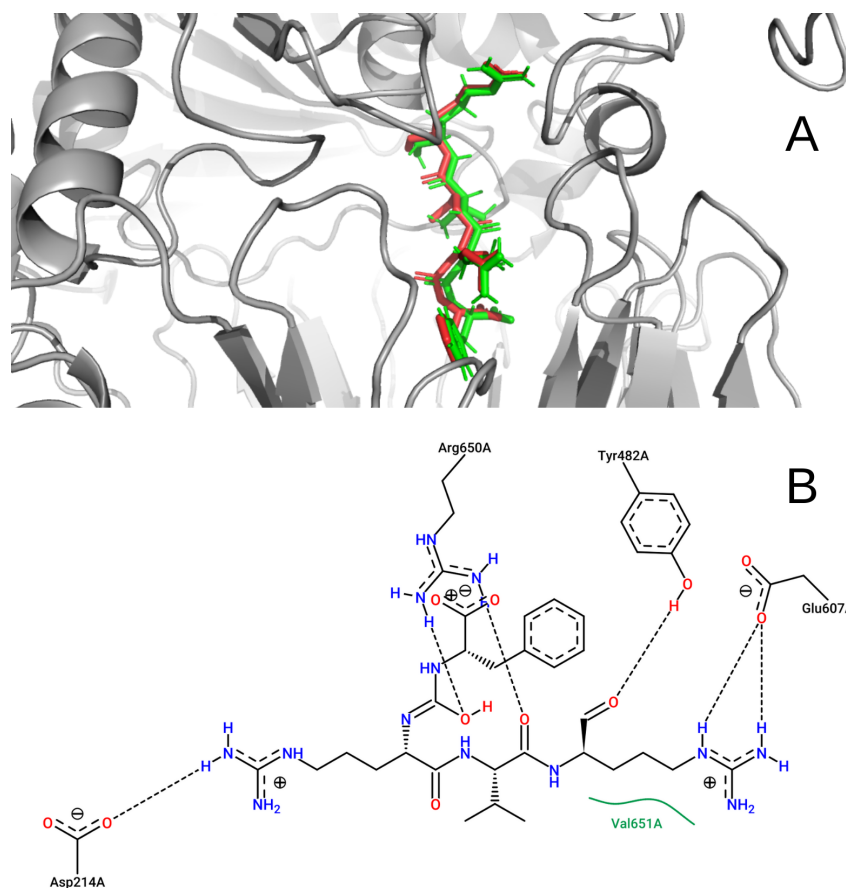


Figura 5. Estrutura 3D do sítio OPBTc com antipaína, na validação por *redocking*. (A) Em vermelho a posição da antipaína na estrutura cristalográfica (PDB ID:4bp9) e em verde a posição do inibidor após o *docking*. (B) Interações intermoleculares do *docking* da antipaína com a OPBTb. As interações com resíduos relevantes são Arg650, Tyr482, Glu607, Asp214 e Val651. Fonte: (SCHÖNING-STIERAND et al., 2020; SCHRÖDINGER, 2015).

O sítio ativo da OPBTb é formado pela tríade catalítica Ser563, Asp648, His683 e outros resíduos também fazem parte da interação da tríade com o substrato como Asp214, Tyr482, Arg 650 e Glu607 assim como determinado pelo servidor PoseView (STIERAND; MAASS; RAREY, 2006; STIERAND; RAREY, 2010). Porém, sabe-se que a ligação da antipaína com o resíduo Ser563 é do tipo

covalente e não é representada pelo servidor *PoseView* devido a limitações do servidor, por isso não aparecerá nas figuras de diagrama 2D.

Foi realizado também o *docking* da antipaína com o modelo da OPBTc e obteve pontuação de $-7,78 \text{ Kcal}\cdot\text{mol}^{-1}$. No entanto, após sobreposição da OPBTc-antipaína e OPBTb-antipaína, pode-se verificar que a geometria e posição da antipaína no sítio não ficou semelhante (Figura 6A) e, conseqüentemente, a interação com o sítio ativo. As interações pertinentes são as ligações de hidrogênio da antipaína com os resíduos Glu606, Asn2012 e Lys206 da OPBTc, assim como interações hidrofóbicas ou dipolo-dipolo induzido Arg649 e Leu602 (Figura 6B).

Por mais que alguns resíduos do sítio de interação da OPBTc com a antipaína sejam conservados em relação à OPBTb, como por exemplo, Arg649 (OPBTc) e Arg650 (OPBTb), o tipo de interação que eles fazem com a antipaína não são os mesmos. O resíduo Arg650 faz duas ligações de hidrogênio, uma na cadeia principal e outra na cadeia secundária da antipaína (Figura 5B). O resíduo Arg649 faz uma interação hidrofóbica com o anel aromático da cadeia secundária da antipaína (Figura 6B).

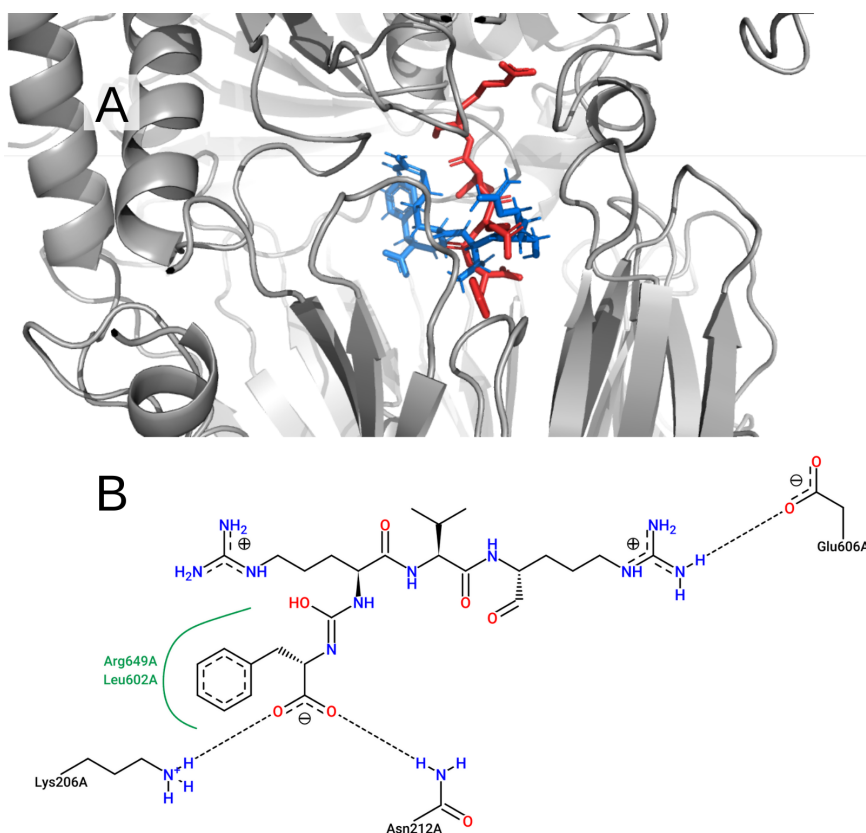


Figura 6. *Docking* da antipaína com a OPBTc. (A) Vermelho: estrutura da antipaína proveniente da OPBTb cristalizada (PDB id:4bp9); Azul: estrutura da antipaína no sítio de ligação à OPBTc. (B) Ligações intermoleculares do *docking* da antipaína com a OPBTc. As interações com resíduos relevantes são Glu606, Asn212, Lys206, Arg649, Leu602. Fonte: modificado pelo autor (SCHÖNING-STIERAND et al., 2020; SCHRÖDINGER, 2015).

A próxima etapa consistiu em realizar o *docking* da OPBTc com os compostos disponíveis na biblioteca *Serine Proteases Inhibitors Library* (ChemDiv). A OPBTc e os compostos foram previamente submetidos a uma preparação estrutural usando as ferramentas Protein Preparation Wizard (SASTRY et al., 2013) e LigPrep (SCHRÖDINGER, 2015). A preparação teve como objetivo simular pH 8,0, pH ótimo para atividade catalítica da OPBTc (MOTTA, 2012). A seleção das coordenadas para ligação dos compostos foi baseada no sítio catalítico da enzima (Ser563, Asp648, His683) que corresponde ao sítio de interação com a antipaína também.

A biblioteca *Serine Proteases Inhibitors Library* (ChemDiv) possui 31.972 compostos e, após a preparação, 32.552 conformações foram obtidas. As pontuações provenientes do experimento de *docking* abrangem de -10,79 até -6,18 Kcal·mol⁻¹. Como a pontuação do parâmetro comparativo (OPBTb + antipaína) foi -8,06 Kcal·mol⁻¹, os compostos com pontuações maiores foram desconsiderados, restando 26.387. Para melhor filtrar a população de ligantes, estabelecemos pontuação máxima de -9,0 Kcal·mol⁻¹, reduzindo os compostos selecionados para 6.004 moléculas.

O próximo filtro utilizado foi o de eficiência dos ligantes. Esse passo tem o propósito de normalizar a afinidade de ligação de um ligante em relação ao número de átomos pesados (não-hidrogênio) (ABAD-ZAPATERO, 2007). Os valores do parâmetro de eficiência de ligantes (LE) ideais (aceitos para fármacos) são acima de 0,3 Kcal · mol⁻¹ · átomo não-hidrogênio⁻¹ (ABAD-ZAPATERO, 2007; KESERÜ; MAKARA, 2009) e, portanto, os ligantes com valores menores que o estipulado foram desconsiderados, filtrando-se assim 2.504 compostos.

Ainda com o objetivo de priorizar os compostos mais promissores, esses 2.504 compostos foram então submetidos ao filtro dos modelos de relação estrutura-atividade quantitativa (QSAR) fenotípico, construídos utilizando dados do *T. cruzi* em ensaios de 72 h. O filtro QSAR fenotípico para *T. cruzi* filtrou 213

compostos com valores de probabilidade de atividade frente a *T. cruzi* acima de 0,6. Esta última etapa foi realizada em colaboração com o LabMol, UFG.

Os 213 compostos foram então submetidos a cálculos no programa DataWarrior utilizando o descritor SkelSpheres para agrupar estruturas com similaridade estrutural (*cluster compounds*) com pontuação maior ou igual a 0,8. Os compostos foram agregados em 66 grupos e filtrou-se somente aqueles que o programa considera como estrutura representativa. Neste caso foram 35 compostos.

A inspeção visual dos 35 compostos restantes foi realizada de forma qualitativa com auxílio das ferramentas de visualização do PyMol. Foram adotados critérios qualitativos empíricos como a boa sobreposição do respectivo ligante com o complexo OPBTc-antipaina e a proximidade dos átomos dos ligantes com os resíduos de importância do sítio ativo (MCLUSKEY et al., 2010), totalizando 10 compostos. Estes critérios foram utilizados para determinar quais compostos seriam promissores e teriam suas interações intermoleculares descritas (Tabela 2).

Tabela 2: Compostos priorizados e suas respectivas pontuações de *docking*, eficiência do ligante e Probabilidade de atividade frente a *T. cruzi*.

Composto	Nome do composto	Docking score (kcal·mol ⁻¹)	Eficiência ligante	do Probabilidade de atividade frente a <i>T. cruzi</i>
1	4-[3-(3-METHOXYPHENYL)-1H-PYRAZOL-5-YL]-N-[1-(3-METHOXYPHENYL)ETHYL]BENZAMIDE	-10,30	0,32	0,93
2	N-{2-[(3-CHLOROPHENYL)CARBAMOYL]-1-BENZOFURAN-3-YL}-5-PHENYL-1,2-OXAZOLE-3-CARBOXAMIDE	-10,27	0,31	0,93
3	N-[(Z)-N'-(5-CHLORO-2-METHOXYPHENYL)-N-[(PYRIDIN-3-YL)METHYL]CARBAMIMIDOYL]-2H-1,3-BENZODIOXOLE-5-CARBOXAMIDE	-10,22	0,32	0,92
4	N-[(E)-N'-(5-CHLORO-2-METHYLPHENYL)-N-[(OXOLAN-2-YL)METHYL]CARBAMIMIDOYL]-3-ACETAMIDO-4-METHYLBENZAMIDE	-10,18	0,32	0,91
6	1-[4-(2-METHYLPROPANAMIDO)PHENYL]-N-[(PYRIDIN-2-YL)METHYL]CYCLOHEXANE-1-CARBOXAMIDE	-10,12	0,36	0,90
17	2-(1H-INDOL-3-YL)-N-[2-OXO-1-PHENYL-2-(1,2,3,4-TETRAHYDROQUINOLIN-1-YL)ETHYL]ACETAMIDE	-9,95	0,31	0,83

19	2-(4-[(4-METHOXYPHENYL)METHYL]AMINO)PHENYL)-N-[(3-(TRIFLUOROMETHYL)PHENYL)METHYL]ACETAMIDE	-9,91	0,31	0,80
22	(E)-N-(4-ETHOXYPHENYL)-N'-[(4-ETHOXYPHENYL)METHYL]-N''-METHANESULFONYLGUANIDINE	-9,89	0,36	0,79
27	N-(1-PHENYLETHYL)-2-[(4-PROPANAMIDOPHENYL)METHYL]-1H-1,3-BENZODIAZOLE-5-CARBOXAMIDE	-9,86	0,30	0,76
34	3-(4-CHLOROPHENYL)-N-[(4-FLUOROPHENYL)METHYL]-1-(2-METHOXYBENZOYL)-1H-1,2,4-TRIAZOL-5-AMINE	-9,84	0,31	0,74

Fonte: elaborado pelo autor.

Com o auxílio dos diagramas 2D gerados pelo PoseView do servidor Proteinplus (SCHÖNING-STIERAND et al., 2020) é possível visualizar interações tridimensionais originalmente complexas em um plano bidimensional. Na inspeção visual das poses de *docking* dos ligantes promissores observou-se as ligações de hidrogênio realizadas pelos heteroátomos dos ligantes, contatos hidrofóbicos dos resíduos com segmentos hidrofóbicos do ligante, interações dos anéis aromáticos do composto realizando pi-stacking com os resíduos próximos da proteína.

Para o sítio da antipaína da OPBTc as interações com os compostos que se reproduzem no *docking* dos ligantes está o composto 1 interagindo com a mesma ligação de hidrogênio em Glu606 (Figura 7). O composto 3 somente interage de forma hidrofóbica em Leu602 (Figura 7). Os compostos 4 e 34 possuem interações apolares em Arg649 e Leu602 de forma semelhante em uma estrutura aromática. Os compostos 6, 7 e 19 se assemelham na interação apolar Leu602, somente (Figura 8). Os compostos 2, 22 e 27 não produzem nenhuma das interações da antipaína (Figura 7, 8 e 9).

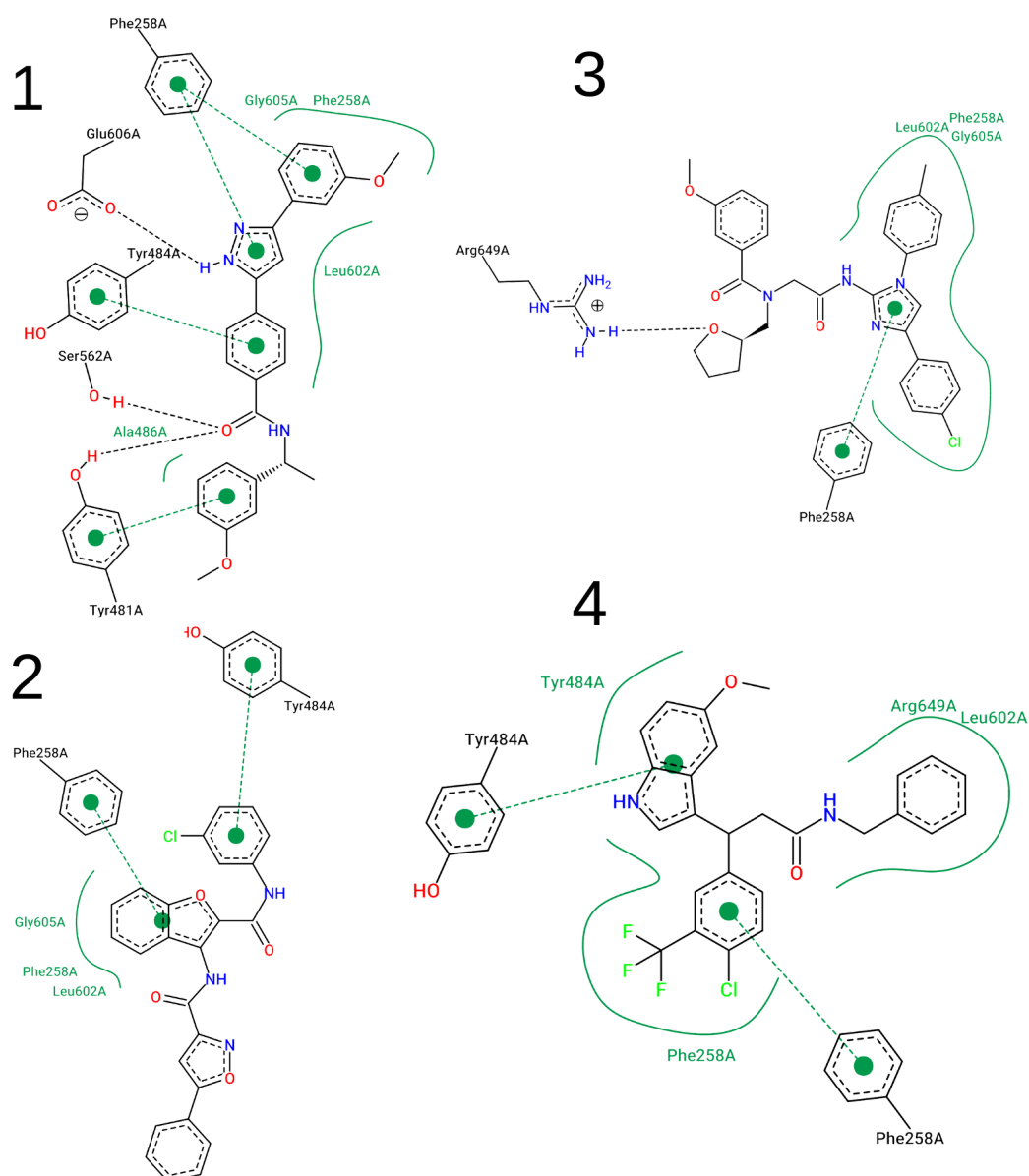


Figura 7. Diagrama 2D de interações intermoleculares dos compostos 1, 2, 3 e 4 com a OPBTc. Destaque para interações pi-stacking com Phe258 presente em todos os compostos.

Por motivo de comparação, semelhantemente com as interações do sítio da antipaína da OPBTb as interações no composto 1 se assemelham em Glu606, sendo esta ligação de hidrogênio. Os compostos 2, 6, 19 e 17 (Figura 7, 8 e 9) não reproduzem nenhuma das interações da antipaína-OPBTb. O composto 3 (Figura 7) reproduz a ligação de hidrogênio em Arg649 (Figura 7). O composto 4 possui interações apolares somente, sendo uma delas Arg649 (Figura 7). O composto 22 interage com a proteína em Tyr481 como ligação de hidrogênio. Os compostos 27 e 34 não reproduzem interações, porém passam a interagir de forma apolar com os resíduos Tyr481 e Arg649 (Figura 9).

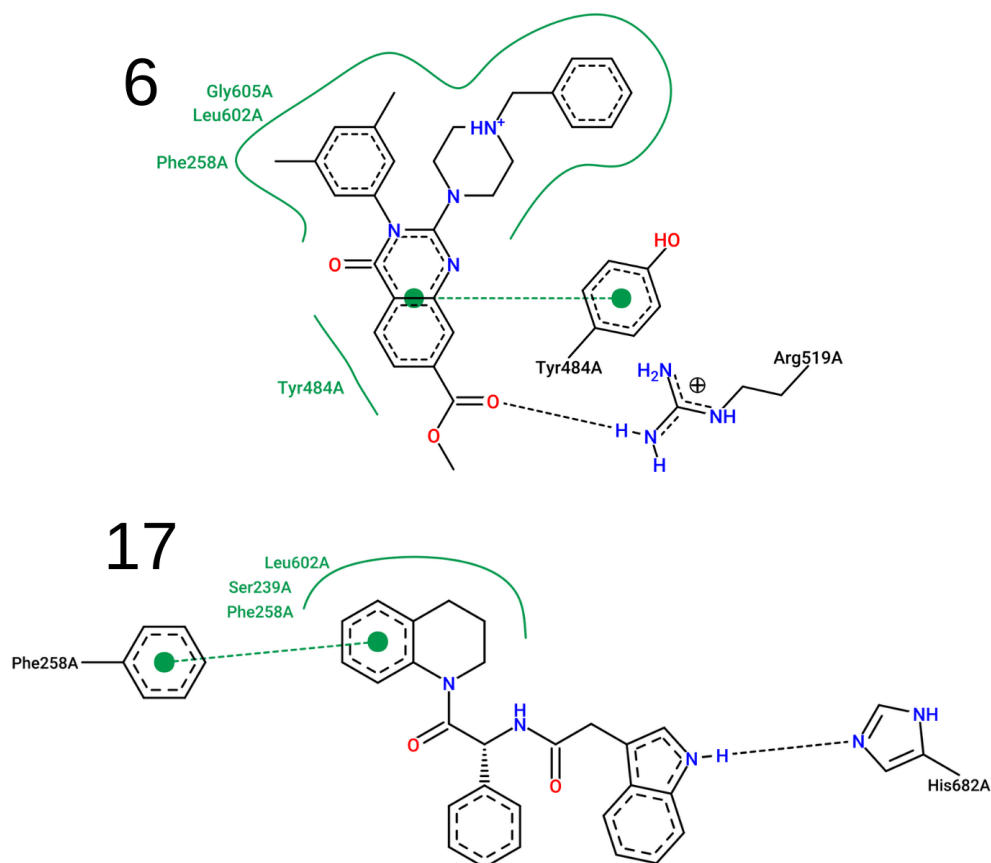


Figura 8. Diagrama 2D de interações intermoleculares dos compostos 6 e 17 com a OPBTc. Destaque para interações ligação de hidrogênio do composto 17 com o resíduo His682 da OPBTc e interações hidrofóbicas com Phe258.

As interações que não são representadas no docking da antipaína com OPBTc ou OPBTb são consistentes entre os ligantes. As interações pi-stacking em Phe258 se repetem nos compostos 1, 2, 3, 4, 17 e 22 (Fig. 7, 8 e 9). A interação pi-stacking ou contato hidrofóbico em Tyr484 ocorre em 1, 2, 4, 6, 19, 22, 27 e 34 (Fig. 7, 8 e 9). A presença destas interações apolares podem reforçar a possível atividade dos ligantes como inibidores.

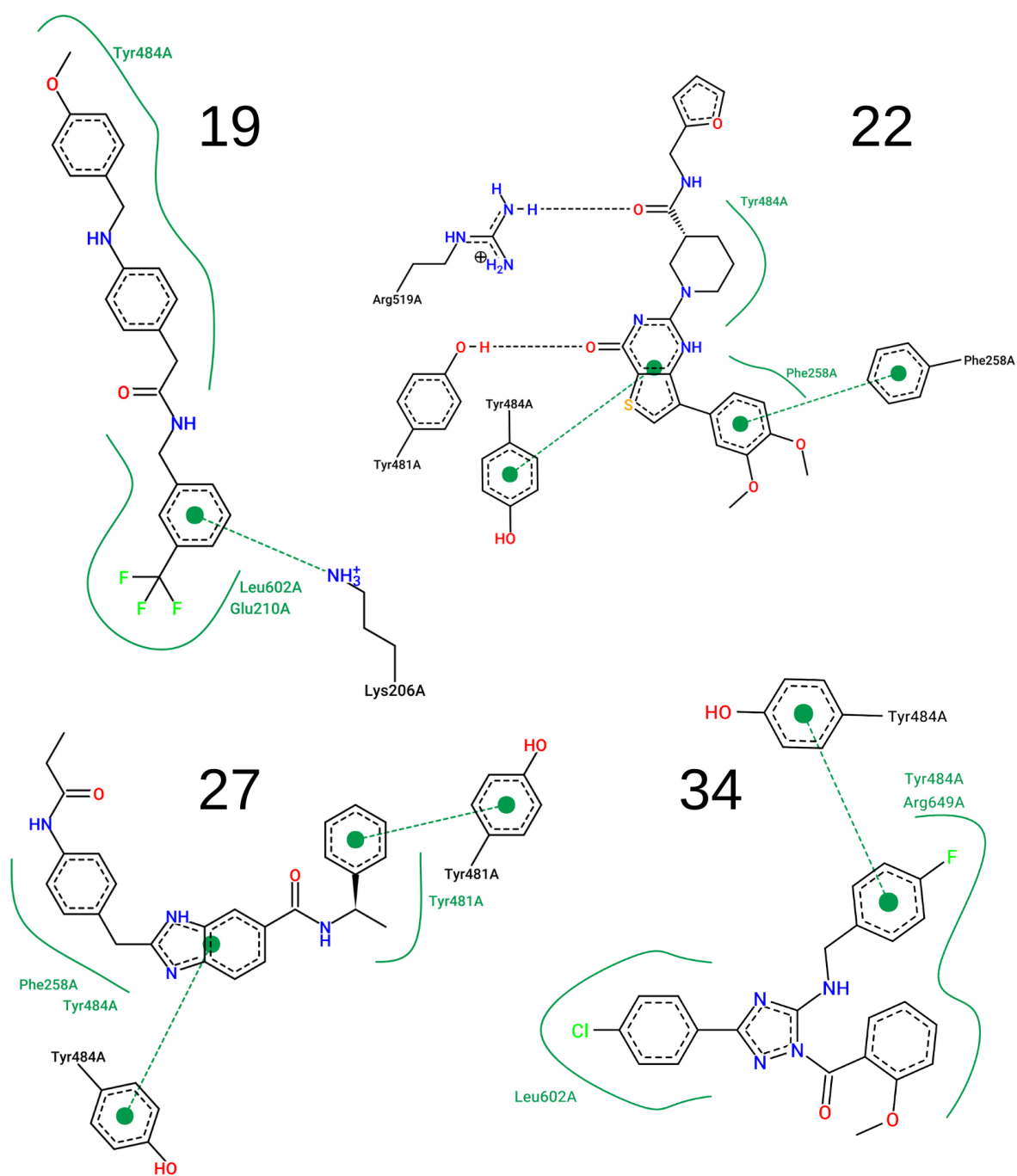


Figura 9. Diagrama 2D de interações intermoleculares dos compostos 19, 22, 27 e 34 com a OPBTc. Destaque para diversas interações hidrofóbicas pi-stacking entre os compostos aromáticos.

Conclusão

Neste trabalho utilizamos diferentes abordagens computacionais modernas e integradas para elucidar a estrutura tridimensional da OPBTc e priorizar compostos potenciais inibidores da enzima, auxiliando assim no descobrimento de novos fármacos no campo de pesquisa de DTN. Os modelos de estrutura 3D da OPBTc gerados foram comparados e analisados quanto a qualidade estrutural. Obtivemos bons modelos para estudo da função e busca por inibidores. O melhor modelo foi a estrutura obtida pelo AlphaFold Q4CW30 e foi escolhida para realizar o *docking* da biblioteca de compostos *Serine Proteases Inhibitors Library* (ChemDiv). Através da integração das abordagens de *docking molecular*, análise de eficiência de ligante, similaridade estrutural e predição de atividade em *T. cruzi*, foi possível priorizar 35 compostos como potenciais inibidores de OPBTc e com atividade tripanocida. Foi possível observar que os compostos se ligam no sítio catalítico da OPBTc, próximo à tríade catalítica e apresentam ligações de hidrogênio e interações hidrofóbicas, com resíduos semelhantes aos observados na antipainá - OPBTb. As etapas futuras deste trabalho incluem a validação experimental desses compostos, através de estudos de atividade enzimática e atividade no parasito.

Bibliografia

- ABAD-ZAPATERO, C. Ligand efficiency indices for effective drug discovery. **Expert opinion on drug discovery**, v. 2, n. 4, p. 469–488, 1 Apr. 2007.
- BERMUDEZ, J. et al. Current drug therapy and pharmaceutical challenges for Chagas disease. **Acta Tropica**, v. 156, p. 1–16, Apr. 2016.
- BURLEIGH, B. A.; ANDREWS, N. W. A 120-kDa alkaline peptidase from *Trypanosoma cruzi* is involved in the generation of a novel Ca(2+)-signaling factor for mammalian cells. **The Journal of Biological Chemistry**, v. 270, n. 10, p. 5172–5180, 10 Mar. 1995.
- BURLEIGH, B. A. et al. A cytosolic serine endopeptidase from *Trypanosoma cruzi* is required for the generation of Ca²⁺ signaling in mammalian cells. **The Journal of Cell Biology**, v. 136, n. 3, p. 609–620, 10 Feb. 1997.
- CANNING, P. et al. Crystal structures of *Trypanosoma brucei* oligopeptidase B broaden the paradigm of catalytic regulation in prolyl oligopeptidase family enzymes. **Plos One**, v. 8, n. 11, p. e79349, 12 Nov. 2013.
- CHEN, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica. Sect. D, Biological Crystallography**, v. 66, n. Pt 1, p. 12–21, Jan. 2010.
- COETZER, T. H. T.; GOLDRING, J. P. D.; HUSON, L. E. J. Oligopeptidase B: a processing peptidase involved in pathogenesis. **Biochimie**, v. 90, n. 2, p. 336–344, Feb. 2008.
- DAVIS, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. **Nucleic Acids Research**, v. 35, n. Web Server issue, p. 375–383, Jul. 2007.
- DE MAGALHÃES, C. S. et al. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. **Information sciences**, v. 289, n. 289, p. 206–224, Dec. 2014.
- DELEN, D.; HELFRICH, S.; SILIPO, R. **KNIME analytics platform for visual data science and business analytics teaching** Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. **Anais...** In: SIGCSE '21: THE 52ND ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION. New York, NY, USA: ACM, 3 Mar. 2021
- DENG, H.; JIA, Y.; ZHANG, Y. Protein structure prediction (Author manuscript). **Int J Mod Phys B**, v. 32, n. 18, p. 1–18, 2018.
- DIHARCE, J. et al. In Silico Peptide Ligation: Iterative Residue Docking and Linking as a New Approach to Predict Protein-Peptide Interactions. **Molecules (Basel, Switzerland)**, v. 24, n. 7, 5 Apr. 2019.
- 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, 19 Oct. 2006.
- GAULTON, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, v. 40, n. Database issue, p. D1100-7, Jan. 2012.
- GÉRCZEI, T.; KESERÜ, G. M.; NÁRAY-SZABÓ, G. Construction of a 3D model of oligopeptidase B, a potential processing enzyme in prokaryotes. **Journal of molecular graphics & modelling**, v. 18, n. 1, p. 7–17, 57, Feb. 2000.

- HAYES, J. M.; STEIN, M.; WEISER, J. Accurate Calculations of Ligand Binding Free Energies: Chiral Separation with Enantioselective Receptors. **The Journal of Physical Chemistry A**, v. 108, n. 16, p. 3572–3580, 1 Apr. 2004.
- HEMMIGE, V.; TANOWITZ, H.; SETHI, A. Trypanosoma cruzi infection: a review with emphasis on cutaneous manifestations. **International Journal of Dermatology**, v. 51, n. 5, p. 501–508, May 2012.
- ILLENBERGER, D. et al. Specificity and structural requirements of phospholipase C-beta stimulation by Rho GTPases versus G protein beta gamma dimers. **The Journal of Biological Chemistry**, v. 278, n. 5, p. 3006–3014, 31 Jan. 2003.
- JANSEN, A. M.; XAVIER, S. C. DAS C.; ROQUE, A. L. R. Trypanosoma cruzi transmission in the wild and its most important reservoir hosts in Brazil. **Parasites & vectors**, v. 11, n. 1, p. 502, 6 Sep. 2018.
- JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, Aug. 2021.
- 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, Nov. 2015.
- KESERÜ, G. M.; MAKARA, G. M. The influence of lead discovery strategies on the properties of drug candidates. **Nature Reviews. Drug Discovery**, v. 8, n. 3, p. 203–212, Mar. 2009.
- KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. Web Server issue, p. W294-7, Jul. 2012.
- KRATZ, J. M. Drug discovery for chagas disease: A viewpoint. **Acta Tropica**, v. 198, p. 105107, Oct. 2019.
- LEBOWITZ, J.; LEWIS, M. S.; SCHUCK, P. Modern analytical ultracentrifugation in protein science: a tutorial review. **Protein Science**, v. 11, n. 9, p. 2067–2079, Sep. 2002.
- MCLUSKEY, K. et al. Crystal structure of Leishmania major oligopeptidase B gives insight into the enzymatic properties of a trypanosomatid virulence factor. **The Journal of Biological Chemistry**, v. 285, n. 50, p. 39249–39259, 10 Dec. 2010.
- MORTY, R. E. et al. Purification and characterisation of a trypsin-like serine oligopeptidase from Trypanosoma congolense. **Molecular and Biochemical Parasitology**, v. 102, n. 1, p. 145–155, 30 Jul. 1999a.
- MORTY, R. E. et al. Oligopeptidase B from Trypanosoma brucei, a new member of an emerging subgroup of serine oligopeptidases. **The Journal of Biological Chemistry**, v. 274, n. 37, p. 26149–26156, 10 Sep. 1999b.
- MOTTA, F. N. et al. The Trypanosoma cruzi virulence factor oligopeptidase B (OPBTc) assembles into an active and stable dimer. **Plos One**, v. 7, n. 1, p. e30431, 19 Jan. 2012.
- MOTTA, F. N. et al. Oligopeptidase B, a missing enzyme in mammals and a potential drug target for trypanosomatid diseases. **Biochimie**, v. 167, p. 207–216, Dec. 2019.
- RAWLINGS, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. **Nucleic Acids Research**, v. 46, n. D1, p. D624–D632, 4 Jan. 2018.
- RIOS, L. E. et al. Immunity and vaccine development efforts against Trypanosoma cruzi.

Acta Tropica, v. 200, p. 105168, Dec. 2019.

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, Apr. 2010.

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, 23 Feb. 2015.

SANTANA, J. M. et al. Purification and characterization of a new 120 kDa alkaline proteinase of *Trypanosoma cruzi*. **Biochemical and Biophysical Research Communications**, v. 187, n. 3, p. 1466–1473, 30 Sep. 1992.

SASTRY, G. M. 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, Mar. 2013.

SCHÖNING-STIERAND, K. et al. ProteinsPlus: interactive analysis of protein-ligand binding interfaces. **Nucleic Acids Research**, v. 48, n. W1, p. W48–W53, 2 Jul. 2020.

SCHRÖDINGER, L. L. C. The PyMOL Molecular Graphics System, Version~1.8. 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, 15 Jul. 2006.

STIERAND, K.; RAREY, M. Drawing the PDB: Protein-Ligand Complexes in Two Dimensions. **ACS Medicinal Chemistry Letters**, v. 1, n. 9, p. 540–545, 9 Dec. 2010.

WATERHOUSE, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. W1, p. W296–W303, 2 Jul. 2018.

YANG, J.; ROY, A.; ZHANG, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. **Bioinformatics**, v. 29, n. 20, p. 2588–2595, 15 Oct. 2013.

YANG, J. et al. The I-TASSER Suite: protein structure and function prediction. **Nature Methods**, v. 12, n. 1, p. 7–8, Jan. 2015.

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, 21 Jan. 2020.

Anexo 01 - Sequência primária

OPBTc (MER0002611):

MMKCGPIAAPKDHEVVFYVEGENRGNNAMNPPRRRNDPLFWLRDDSRKNPEVIAHLKLEQAYFEE
 RTADIKDFSESIKEYISHIKETDTSAPYRYDGYTTYTREVEGLSYKIHCRVPLGMTPGKSEDEQVILDE
 NKLAEGKSFCMVHEVKPARPDHNLVAYSVDYVGNEMYTIRFTDNAVTDVVEGTNGQILWGNASCFF
 YITKDAERDYRIWRHIIGRPQSEDVCLYTENDPLFSTAMAIISGDGNTLIIGSGSFETTESHLLDLRKG
 GHTTLEMVRPREKGVRYDVELHGTETLLILTNDKNCMNGKVV SATRNTPSDWENVIVPHSEDVFIGEI
 GVFAKFVVLSGRRGGLTRVWAMPVGS DGLFRSGAFLQEVSFDEPVFTAAPVFSHMKMYDTETLRVSY
 TSMSTPTTWFDLHVNGTRTIVKVREVLGNF DAKNYTCRRLFATAPDRTKIPLSIVHDVSLDMSKPHPT
 VLYAYGSYGACVEPEFSVKYLPYLDRGVIVIAHV RGGGEMGRAWYEVGAKYLTKRNTFSDFIACA
 EYLIEIGLTTPSQLACEGRSAGLLIGAVLNMRPDLFRVALAGVPFVDVMTTMCDP SIPLTTGEWEE
 WGNPNEYKFFDYMN SYSPVDNVRAQDYPHLMIQAGLHDPVAYWEPAKWASKLRALKTDSNEVLLK
 MDLES GHFSASDRYRYWREMSFQQAFVLKHLNARTLLRR

OPBTb (PDB 4bp9):

MKCGPIATPKDHEVVFYVEGENRGNNAMNPPRRRNDPLFWLRDDSRKNPEVIAHLKLEQAYFEERT
 ADIKDFSETIFKEYISHIKETDISAPYRYDGYTTYTREVEGLSYKIHCRVPLGMTPGKSEDEQVIL
 DENKLAEGKPFVCMVHEVKPAPPDHNLVAYSVDYIGNEMYTIRFTDNAVTDVVEGTNGQILWGN
 ASCFFYMTKDAAERDYKIWRHIIGRPQSEDVCLYTENDLLFSTAMAIISGDGSTLIIGSGSFETTES
 QLLDLRKGNGHTTLEMVRPREKGVRYDVELHGTENLLILTNDKNCMNGKVV LATRNTPSDWENI
 IVPHSEDVFIGEIGVFAKFAVLSGRRGGLTRVWTMPVGS DGLFRSGAFVQEVSFDEPVFTAAPV
 FSHMKMYDTETLRVSYTSMSTPTTWFDLHVNGTRTIVKVREVLGNF DAKNYTCRRLFATAPDR
 TKIPLSIVHDVSLDMSKPHPTVLYAYGSYGVCVEPEFSVKYLPYLDRGVIVIAHV RGGGEM
 GRAWYEVGAKYLTKRNTFSDFIACA EYLIEIGLTTPSQLACEGRSAGLLIGAVLNMRPDLFR
 VALAGVPFVDVMTTMCDP SIPLTTGEWEEWGNPNEYKFFDYMN SYSPVDNVRAQDYPHLMIQ
 AGLHDPVAYWEPAKWASKLRALKTDSNEVLLKMDLES GHFSASDRYRYWREMSFQQAFVLKHL
 NARTLLRR