



**UNIVERSIDADE FEDERAL DE GOIÁS
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA
TROPICAL
E SAÚDE PÚBLICA**

VINÍCIUS ALEXANDRE FIAIA COSTA

**Abordagem computacional para a descoberta de novos inibidores
de prolil oligopeptidase 80 de *Trypanosoma cruzi***

**Goiânia
2022**



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE PATOLOGIA TROPICAL E SAÚDE PÚBLICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese

2. Nome completo do autor

VINÍCIUS ALEXANDRE FIAIA COSTA

3. Título do trabalho

Abordagem computacional para a descoberta de novos inibidores de prolil oligopeptidase 80 de *Trypanosoma cruzi*

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a) consulta ao(à) autor(a) e ao(à) orientador(a);
- b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **VINÍCIUS ALEXANDRE FIAIA COSTA, Discente**, em 28/11/2022, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Bruno Junior Neves, Professor do Magistério Superior**, em 28/11/2022, às 16:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2949598** e o código CRC **8D742922**.

Processo: 23070.024581/2022-61
Documento: 4432088



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE PATOLOGIA TROPICAL E SAÚDE PÚBLICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

VINÍCIUS ALEXANDRE FIAIA COSTA

3. Título do trabalho

Abordagem computacional para a descoberta de novos inibidores de prolil oligopeptidase 80 de *Trypanosoma cruzi*

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(a) autor(a) e ao(a) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Vinícius Alexandre Fiaia Costa, Usuário Externo**, em 06/03/2024, às 14:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4432088** e o código CRC **B410001F**.

VINÍCIUS ALEXANDRE FIAIA COSTA

**Abordagem computacional para a descoberta de novos inibidores
de prolil oligopeptidase 80 de *Trypanosoma cruzi***

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Medicina Tropical e Saúde Pública da Universidade Federal de Goiás para obtenção do Título de Mestre em Medicina Tropical e Saúde Pública.

Orientador: Prof. Dr. Bruno Junior Neves
Coorientadora: Profa. Dra. Carolina Horta Andrade

**Goiânia
2022**

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Costa, Vinícius Alexandre Fiaia
Abordagem computacional para a descoberta de novos inibidores de prolil oligopeptidase 80 de Trypanosoma cruzi [manuscrito] / Vinícius Alexandre Fiaia Costa. - 2022.
CIII, 103 f.: il.

Orientador: Prof. Dr. Bruno Junior Neves; co-orientadora Dra. Carolina Horta Andrade.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Patologia Tropical e Saúde Pública (IPTSP), Programa de Pós Graduação em Medicina Tropical e Saúde Pública, Goiânia, 2022.
Bibliografia.
Inclui siglas, abreviaturas, tabelas, lista de figuras, lista de tabelas.
1. Doença de Chagas. 2. CADD. 3. modelagem por homologia. 4. ancoramento molecular. 5. triagem virtual. I. Neves, Bruno Junior, orient. II. Título.

CDU 615.1

**PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL E
SAÚDE PÚBLICA DA UNIVERSIDADE FEDERAL DE GOIÁS**

BANCA EXAMINADORA DA DISSERTAÇÃO DE MESTRADO

Aluno: Vinícius Alexandre Fiaia Costa

Orientador: Prof. Dr. Bruno Junior Neves

Coorientadora: Profa. Dra. Carolina Horta Andrade

Membros:

1. Prof(a). Dr(a). Bruno Junior Neves

2. Prof(a). Dr(a). Helioswilton Sales de Campos

3. Prof(a). Dr(a). Izabela Marques Dourado Bastos Charneau

Data: 31/05/2022



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE PATOLOGIA TROPICAL E SAÚDE PÚBLICA

ATA DE DEFESA DE DISSERTAÇÃO

ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE DISSERTAÇÃO DE VINÍCIUS ALEXANDRE FIAIA COSTA – Aos trinta e um dias do mês de maio do ano de 2022 (31/05/2022), às 13h30min, reuniram-se os componentes da Banca Examinadora: Profs. Drs. **BRUNO JUNIOR NEVES** (UFG), **HELIOSWILTON SALES CAMPOS** (UFG) e **IZABELA MARQUES DOURADO BASTOS CHARNEAU** (UnB), para, sob a presidência do primeiro, e em sessão pública realizada por WEBCONFERÊNCIA, procederem à avaliação da defesa de dissertação intitulada: “**Abordagem computacional para a descoberta de novos inibidores de prolil oligopeptidase 80 de *Trypanosoma cruzi***”, em nível de **MESTRADO**, área de concentração em **CIÊNCIAS BÁSICAS E APLICADAS EM DOENÇAS INFECTOPARASITÁRIAS E SAÚDE PÚBLICA: PARASITOLOGIA**, de autoria de **VINÍCIUS ALEXANDRE FIAIA COSTA**, discente do PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL E SAÚDE PÚBLICA, da Universidade Federal de Goiás. A sessão foi aberta pelo Orientador, Prof. Dr. **BRUNO JUNIOR NEVES**, que fez a apresentação formal dos membros da Banca e orientou o Candidato sobre como utilizar o tempo durante a apresentação de seu trabalho. A palavra a seguir, foi concedida ao autor da dissertação que, em 30 minutos procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca arguiu o Candidato, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se à avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1481/2017 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Medicina Tropical e Saúde Pública. A banca, em sessão secreta, expressou seu Julgamento, considerando o candidato **Aprovado** ou **Reprovado**:

Banca Examinadora	Aprovado / Reprovado
Prof. Dr. Bruno Junior Neves	Aprovado
Prof. Dr. Helioswilton Sales Campos	Aprovado
Profa. Dra. Izabela Marques Dourado Bastos Charneau	Aprovado

Em face do resultado obtido, a Banca Examinadora considerou o candidato **Habilitado**, (**Habilitado** ou **não Habilitado**), cumprindo todos os requisitos para fins de obtenção do título de **MESTRE EM MEDICINA TROPICAL E SAÚDE PÚBLICA**, na área de concentração em **CIÊNCIAS BÁSICAS E APLICADAS EM DOENÇAS INFECTOPARASITÁRIAS E SAÚDE PÚBLICA: PARASITOLOGIA** 16 h e 15 min, a presidência da mesa encerrou esta sessão de defesa de dissertação e para constar eu, HELOÍSA DE SOUSA VIEIRA, secretária do Programa de Pós-Graduação em Medicina Tropical e Saúde Pública lavrei a presente Ata que depois de lida e aprovada, será assinada pelos membros da Banca Examinadora.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Helioswilton Sales De Campos, Professor do Magistério Superior**, em 31/05/2022, às 16:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Bruno Junior Neves, Professor do Magistério Superior**,



em 31/05/2022, às 16:28, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **IZABELA MARQUES DOURADO BASTOS CHARNEAU, Usuário Externo**, em 01/06/2022, às 11:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2912223** e o código CRC **D0331788**.

Referência: Processo nº 23070.024581/2022-61

SEI nº 2912223

*Dedico este trabalho aos meus pais,
Moisés e Viviane, que sempre me
incentivaram a buscar conhecimento.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me concedido a oportunidade de realizar o sonho de me tornar um mestre.

Aos meus pais, Moisés Alexandre Costa e Viviane de Mendonça Fiaia Costa, pelo incentivo e apoio, me guiando sempre pelos melhores caminhos.

Ao professor Dr. Bruno Junior Neves, pela competência, orientação e oportunidade de realizar este trabalho.

À professora Dra. Carolina Horta Andrade, pelos ensinamentos, exemplo de profissionalismo e dedicação.

À professora Dra. Izabela Marques Dourado Bastos Charneau (Departamento de Biologia Celular da Universidade de Brasília), que conduzirá a validação experimental deste trabalho e também pela contribuição como membro da banca de defesa.

Ao professore Dr. Gustavo Henrique Goulart Trossini pela contribuição na banca de qualificação de mestrado.

Ao professor Dr. Helioswilton Sales Campos pelos conselhos e sugestões feitas na banca de qualificação e defesa de mestrado.

A todos os colegas do LabMol por toda ajuda e atenção nos momentos de dificuldade.

À OpenEye Scientific Software e ChemAxon pela licença acadêmica de programas utilizados neste trabalho.

Ao CNPq pela concessão da bolsa de mestrado e pelo financiamento do projeto “Inteligência Artificial para a Descoberta de Novos Candidatos a Fármacos Eficazes contra Tripanossomatídeos”, Chamada Universal 425119/2018-1.

SUMÁRIO

QUADROS, TABELAS, FIGURAS E ANEXOS.....	12
SÍMBOLOS, SIGLAS E ABREVIATURAS	17
RESUMO	20
ABSTRACT.....	21
1. INTRODUÇÃO	21
1.1. Doença de Chagas.....	21
1.1.1. Etiologia e distribuição.....	22
1.1.2. Ciclo Biológico	22
1.1.3. Tratamento Farmacológico.....	25
1.2. Alvos moleculares para o planejamento de novos fármacos	27
1.2.1. Prolil oligopeptidase 80 de <i>T. cruzi</i>	30
1.3. Planejamento e descoberta de novos fármacos.....	34
1.3.1. Planejamento de fármacos auxiliado por computador	37
1.3.2. Predição de estruturas de proteínas.....	38
1.3.3. Ancoramento molecular	41
1.3.4. Modelos de forma e volume molecular	45
1.3.5. Triagem Virtual.....	46
2. JUSTIFICATIVA	47
3. OBJETIVOS	48
3.1. Objetivo geral.....	48
3.2. Objetivos específicos	48
4. MÉTODOS	49
4.1. Métodos computacionais	49
4.1.1. Predição da estrutura 3D	49
4.1.2. Refinamento estrutural e validação da POPTc80	49
4.1.3. Análise estrutural da POPTc80.....	49
4.1.4. Predição dos sítios de ligação na POPTc80	50
4.1.5. Avaliação do nível de conservação dos aminoácidos da POPTc80	51
4.1.6. Compilamento e preparo do conjunto de validação	52
4.1.7. Preparo do conjunto de validação	53
4.1.8. Preparo da proteína	53
4.1.9. Ancoramento molecular	53
4.1.10. Modelos <i>shape-based</i>	54

4.1.11.	Validação estatística dos modelos gerados	56
4.1.12.	Triagem Virtual.....	58
4.2.	Métodos experimentais.....	58
4.2.1.	Ensaio enzimáticos	58
5.	RESULTADOS	59
5.1.	Predição da estrutura 3D da POPTc80.....	59
5.2.	Análise estrutural da POPTc80.....	61
5.3.	Validação do ancoramento molecular	65
5.4.	Modelos Shape-Based.....	70
5.5.	Triagem Virtual.....	74
	DISCUSSÃO	82
	CONCLUSÕES	87
	REFERÊNCIAS	88

QUADROS, TABELAS, FIGURAS E ANEXOS

- Figura 1-** Distribuição global da doença de Chagas, baseada em estimativas oficiais, no período de 2006-2015 (Adaptado de WORLD HEALTH ORGANIZATION, 2020). 22
- Figura 2 -** Ciclo biológico de *Trypanosoma cruzi*. (1) O inseto triatomíneo realiza repasto sanguíneo e libera tripomastigotas metacíclicas nas fezes, os quais adentram o hospedeiro vertebrado através de abertura no local da picada ou mucosas intactas; (2) dentro do hospedeiro vertebrado, os tripomastigotas invadem células e se diferenciam em amastigotas intracelulares; (3) As amastigotas se multiplicam dentro das células; (4) Em seguida, as amastigotas intracelulares se diferenciam em tripomastigotas, rompem as células e são liberadas na corrente sanguínea, podendo adentrar novas células, se diferenciar novamente em amastigotas, e se replicarem; (5) O triatomíneo, ao realizar repasto sanguíneo, ingere tripomastigotas circulantes; (6) As tripomastigotas ingeridos se diferenciam em epimastigotas no intestino médio do vetor; (7) as epimastigotas se multiplicam no intestino médio do vetor; (8) Ao final deste processo, as epimastigotas se diferenciam em tripomastigotas metacíclicas no intestino posterior (Adaptado de (CDC, 2021). 23
- Figura 3 -** Formas morfológicas do *T. cruzi*. Forma amastigota (a), epimastigota (b), tripomastigota circulante (c) e tripomastigota metacíclica (d) (FIOCRUZ, 2017)... 24
- Figura 4 –** Organelas celulares das formas Amastigota (a), epimastigota (b) e tripomastigota (c) (TEIXEIRA et al., 2012). 25
- Figura 5 -** Estrutura química do fármaco benznidazol (1), utilizado no tratamento da doença de Chagas..... 25
- Figura 6 –** Representação esquemática do modo de ação dos principais fármacos com atividade tripanocida. As caixas verdes representam fármacos atualmente usados para tratamento (exceto Nifurtimox), as caixas amarelas representam medicamentos em ensaios clínicos e as caixas vermelhas representam medicamentos experimentais. (a) Os bifosfonatos inibem a farnesil pirofosfato sintase, que reduz os níveis de esteróis e outros compostos poliisoprenoides essenciais, afetando a viabilidade das células. (b) Nirfurtimox (NFX) e benznidazol (BZN) são reduzidos pela nitrorredutase do parasito, resultando na produção de espécies de oxigênio reativo (ROS), que danificam diretamente as células do parasito. A tripanotiona redutase ajuda a aliviar o estresse

oxidativo, e os inibidores dessas enzimas, como a tioridazina e a sulfoximina butionina, aumentam a quantidade de ROS no espaço intracelular. (c) O ergosteral biossintético é essencial para a sobrevivência do parasito. O bloqueio dessa via leva à perda da viabilidade celular por meio da depleção de esteróis essenciais e acúmulo de intermediários tóxicos. Inibidores de ergosterol e análogos de fenarimol têm como alvo a lanosterol C14 demetilase, e amiodarona e dronedarona parcialmente inibe a oxidosqualeno ciclase. Os inibidores de fosfolipídios bloqueiam a síntese de esteróis, inibem a síntese de novo de fosfolipídios pela via de Greenberg e inibem enzimas de transdução de sinal como a fosfatidilinositol fosfolipase C. (d) Amiodarona e dronedarona liberam Ca^{2+} das mitocôndrias e acidocalcisomas, o que aumenta os níveis de Ca^{2+} no espaço citoplasmático e compromete a sobrevivência celular. (e) Cruzipaína (CZ) é tipicamente localizada no complexo de Golgi, bolsa flagelar e glicosomas e é uma cisteína protease essencial envolvida na diferenciação do parasito, invasão celular, multiplicação e evasão imune. Os inibidores da cruzipaína alteram o complexo de Golgi, devido ao acúmulo de precursores da cruzipaína (SALES JUNIOR et al., 2017)..... 27

Figura 7 – Compostos com atividade inibitória para POPTc80 (BAL et al., 2003; JOYEAU et al., 2000; VENDEVILLE et al., 1999a). 31

Figura 8 - O mecanismo proteolítico utilizado pelas serina proteases “clássicas” Ser/His/Asp. As etapas do mecanismo catalítico de protease de serina Ser/His/Asp "clássico" incluem: complexo de Michaelis (a), intermediário tetraédrico 1 (b), intermediário acil-enzima (c) e intermediário tetraédrico 2 (d)(EKICI; PAETZEL; DALBEY, 2008). 33

Figura 9 - Modelos estruturais de Prolil Oligopeptidase. Modelo estrutural da POPTc80 (a) gerado pelo AlphaFold, a POPTc80 é dividida em um domínio α/β hidrolase catalítico (salmão) composto por α -hélices e folhas β , e um domínio β -propeller (azul claro) formado somente por formado somente por folhas β . Estrutura da POP de *Sphingomonas capsulata* aberta (b)(SHAN; MATHEWS; KHOSLA, 2005). 34

Figura 10 – Etapas envolvidas na pesquisa e desenvolvimento de fármacos (Cortesia do Prof. Dr. Bruno Junior Neves). 35

Figura 11 – Abordagens para predição de estrutura 3D. Modelagem por homologia, Predição por enovelamento e Predição *Ab initio*. 39

Figura 12 – Fluxograma de um típico ancoramento molecular. 42

Figura 13 - Representação da geração de modelos baseados na forma e volume molecular no programa ROCS (OPENEYE SCIENTIFIC SOFTWARE, 2020).	45
Figura 14 – Etapas envolvidas em uma típica triagem virtual.	46
Figura 15 - Diagrama esquemático relacionando o desempenho do acoplamento molecular à curva ROC. A classificação a taxa na curva <i>a</i> é melhor do que <i>b</i> , <i>c</i> e <i>e</i> , respectivamente. A curva ROC para um encaixe aleatório a classificação é representada como uma li linha pontilhada em <i>d</i> (Adaptado de NEVES et al., 2021).	56
Figura 16 – Estrutura 3D da POPTc80 gerada pelos servidores SWISS-MODEL (a), I-TASSER (b) e Alphafold (c). Domínio α/β hidrolase e domínio β -propeller representado em salmão e azul claro, respectivamente.	59
Figura 17 – Análise de PCA da POPTc80 e suas flutuações. Análise de PCA do modelo gerado pelo AlphaFold. Comparação entre PC1 e PC2 (a); PC1 e PC3 (b) e a proporção da variância relacionada à classificação de autovalor (c) e gráfico de flutuações (d). Cada ponto representa uma estrutura e a cor do ponto indica o ID de cluster de um cluster conformacional.	62
Figura 18 – Predição de sítios de ligação nas conformações Conf09, Conf15 e Conf20 de POPTc80 utilizando os servidores FTSite (a) e análise de conservação dos resíduos de aminoácido baseada no servidor ConSurf (b).	63
Figura 19 – Estrutura 2D dos inibidores co-cristalizados de prolil oligopeptidases e seus valores de inibição biológico.	64
Figura 20 – Modelo hipotético do modo de ligação de inibidores de POPTc80 (a) e o posicionamento da <i>Query</i> 4 em relação ao sítio da POPTc80 (b).	65
Figura 21 – Diagrama 3D das <i>queries</i> 1, 2 e 3 dos modelos <i>Shape-based</i> , sendo as cores: doador (azul tracejado), aceitador (vermelho tracejado) e anel (verde).	71
Figura 22 – Diagrama 3D das <i>queries</i> 4, 5 e 6 dos modelos <i>Shape-based</i> , sendo as cores: aceitador (vermelho tracejado) e anel (verde).	73
Figura 23 – Etapas de filtragem exploradas na VS de novos inibidores de POPTc80 utilizando a base de dados ChemBridge.	75
Figura 24 – Interações moleculares do <i>hit</i> 7898173. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.	78

Figura 25 - Interações moleculares do <i>hit</i> 5634226. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.	79
Figura 26 - Interações moleculares do <i>hit</i> 7991224. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.	79
Figura 27 - Interações moleculares do <i>hit</i> 7961087. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) em verde tracejado, interações do tipo π -cátion em laranja tracejado. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.	80
Figura 28 - Interações moleculares do <i>hit</i> 7886969. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) em verde tracejado, interações do tipo π -cátion em laranja tracejado. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.	81
Tabela 1 - Conjunto de sondas moleculares utilizadas pelo FTSite.	50
Tabela 2 – Coordenadas dos grids construídos para a POPTc80 em ambos os sítios.	53
Tabela 3 – Compostos utilizados para a criação de modelos <i>Shape-based</i>	55
Tabela 4 – Validação dos modelos 3D gerados pelo SWISS-MODEL, I-TASSER e AlphaFold, respectivamente.	60
Tabela 5 - Validação dos modelos 3D gerados pelo SWISS-MODEL, I-TASSER e AlphaFold após o refinamento no servidor GalaxyWEB.	60
Tabela 6 – Resultado da validação dos protocolos de ancoramento molecular no sítio 1 da enzima POPTc80 utilizando a função de pontuação <i>Docking score</i>	66
Tabela 7 – Resultado da validação dos protocolos de ancoramento molecular no sítio 2 da enzima POPTc80 utilizando a função de pontuação <i>Docking score</i>	66

Tabela 8 - Resultado da validação dos protocolos de ancoramento molecular no sítio 1 da enzima POPTc80 utilizando a função de pontuação <i>Glide Emodel</i>	67
Tabela 9 – Resultado da validação dos protocolos de ancoramento molecular no sítio 2 da enzima POPTc80 utilizando a função de pontuação <i>Glide Emodel</i>	68
Tabela 10 – Resultado da validação dos protocolos de ancoramento molecular no sítio 1 da enzima POPTc80 utilizando a função de pontuação <i>Glide Energy</i>	69
Tabela 11 – Resultado da validação dos protocolos de ancoramento molecular no sítio 2 da enzima POPTc80 utilizando a função de pontuação <i>Glide Energy</i>	69
Tabela 12 - Desempenho estatístico dos modelos de <i>Shape-Based</i> desenvolvidos neste trabalho.....	74
Tabela 13 - <i>Hits</i> virtuais selecionados a partir da VS.	75

SÍMBOLOS, SIGLAS E ABREVIATURAS

2D	Bidimensional
3D	Tridimensional
ACD	Acetamida
ACN	Actonitrila
ACT	Acetona
ADY	Acetaldeído
AMN	Metanamina
ANN	Modelos de rede anisotrópica (do inglês, <i>Anisotropic network models</i>)
AUC	Área sob a curva (do inglês, <i>Area under curve</i>)
BACE1	Beta-secretase 1
BDY	Benzaldeído
BEDROC	Discriminação aprimorada de Boltzmann da curva ROC (do inglês, <i>Boltzmann-Enhanced Discrimination of ROC</i>)
BEN	Benzeno
BUT	Tert-butanol
BZN	Benznidazol
Cα	Carbonos alfa
CADD	Planejamento de fármacos auxiliado por computador (do inglês, <i>Computer Assisted Drug Design</i>)
CHX	Ciclohexano
CA	Cruzipaína
DFO	<i>N,N</i> -dimetilformamida
DME	Dimetil éter
DNDi	<i>Drugs for Neglected Diseases initiative</i>
DTUs	<i>Discrete typing units</i>
DUD	Diretório de decoys úteis (do inglês, <i>directory of useful decoys</i>)
EF	Fator de enriquecimento (do inglês, <i>Enrichment factor</i>)
EN	Rede elástica (do inglês, <i>Elastic network</i>)
EOL	Etanol

ETH	Etano
GHIT	<i>Global Health Innovative Technology Fund</i>
GMQE	Modelo Global de Qualidade Estimada
HTS	Triagem biológica automatizada em grande escala (do inglês, <i>High Throughput Screening</i>)
IC₅₀	Dose capaz de inibir 50% dos parasitos
Ki	Constante de ionização
LBDD	Planejamento baseado no ligante (do inglês, <i>Ligand-Based Drug Design</i>)
LBVS	Triagem virtual baseada no ligante (do inglês, <i>Ligand-based virtual screening</i>)
LogP	Coefficiente de partição octanol/água
LogS	Solubilidade em água
NFX	Nifurtimox
NMR	Ressonância magnética nuclear (do inglês, <i>Nuclear magnetic resonance</i>)
OMS	Organização mundial da saúde
PAINS	<i>Pan-assay interference compounds</i>
PC	Componente principal (do inglês, <i>Principal component</i>)
PCA	Análise de componentes principais (do inglês, <i>Principal component analysis</i>)
P&D	Pesquisa & Desenvolvimento
PDB	<i>Protein Data Bank</i>
PHN	Fenol
POP	Prolil Oligopeptidase
POPTc80	Prolil Oligopeptidase 80 de <i>Trypanosoma cruzi</i>
QSAR	Relações quantitativas entre estrutura e atividade (do inglês, <i>Quantitative Structure-Activity Relationships</i>)
QSQE	Estimativa de qualidade da estrutura quaternária
ROC	<i>Receiver operating characteristic</i>
ROS	Espécies de nitrogênio reativo
SBDD	Planejamento de fármacos baseado na estrutura (do inglês, <i>Structure-Based Drug Design</i>)

SBVS	Triagem virtual baseada na estrutura (do inglês, <i>Structure-based virtual screening</i>)
<i>T. brucei</i>	<i>Trypanosoma brucei</i>
<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>
THS	Isopropanol
UNB	Universidade de Brasília
URE	Ureia
VS	Triagem virtual (do inglês, <i>Virtual Screening</i>)

RESUMO

Título: Abordagem computacional para a descoberta de novos inibidores de prolil oligopeptidase 80 de *Trypanosoma cruzi*

A doença de Chagas é uma doença infecciosa grave causada pelo parasito tripanossomatídeo *Trypanosoma cruzi*. Atualmente, o fármaco benznidazol representa a única opção farmacológica para tratamento da doença. Entretanto, o surgimento de parasitos resistentes, baixa eficácia na fase crônica da doença e número considerável de efeitos adversos do benznidazol tornam premente a descoberta de novos fármacos mais eficazes, seguros e com mecanismo de ação inovador. Neste contexto, o objetivo geral desse trabalho foi identificar inibidores da enzima prolil oligopeptidase 80 (POPTc80), um alvo validado para *Trypanosoma cruzi*, utilizando métodos computacionais baseados na estrutura do alvo (SBDD) e do ligante (LBDD). Inicialmente, a estrutura tridimensional da POPTc80 foi predita utilizando três abordagens diferentes. Após refinamento estrutural e validação, a melhor estrutura 3D, obtida no servidor AlphaFold foi submetida a análise estrutural. Nesta etapa, três conformações representativas da POPTc80 geradas no servidor DynOmics foram selecionadas utilizando Análise de Componentes Principais (PCA). Essas estruturas foram então utilizadas com base estruturais para construção e validação dos protocolos de ancoramento molecular no programa Glide. O ancoramento molecular demonstrou que a conformação 20 resulta na maior taxa de enriquecimento durante triagem de um conjunto de compostos ativos e *decoys*. Em seguida, as poses de ancoramento para inibidores de POPTc80 descritos na literatura foram consideradas *queries* para geração e validação de modelos *shape-based* no programa vROCS. Como resultados, foi possível uma estrutura 3D da POPTc80, foi desenvolvido um protocolo de docking validado e também foi desenvolvido um modelo *shape-based* específico para POPTc80. Ao final deste processo, os melhores protocolos de ancoramento molecular e melhor modelo *shape-based* foram utilizados como filtros computacionais para a triagem virtual da biblioteca ChemBridge, o que possibilitou a obtenção de um conjunto de *hits* putativos que serão validados experimentalmente na enzima POPTc80 em colaboração.

Palavras-chaves: Doença de Chagas, CADD, modelagem por homologia, ancoramento molecular, triagem virtual.

ABSTRACT

Title: Computational approach to the discovery of new *Trypanosoma cruzi* prolyl oligopeptide 80 inhibitors

Chagas disease is a serious infectious disease caused by the trypanosomatid parasite *Trypanosoma cruzi*. Currently, the benznidazole is the only drug available for treating the disease. However, due to the emergence of benznidazole-resistant parasites, low efficacy in the chronic phase and considerable number of adverse effects, the discovery of new drugs more effective, safer and with innovative mechanism of action is imperative. In this context, the general objective of this work was to identify inhibitors of the enzyme prolyl oligopeptidase 80 (POPTc80), a validated target for *Trypanosoma cruzi*, using computational methods based on the structure of the target (SBDD) and ligand (LBDD). Initially, the three-dimensional structure of POPTc80 was predicted using three different approaches. After structural refinement and validation, the best 3D structure, obtained from the AlphaFold server, was submitted to structural analysis. In this step, three representative conformations of POPTc80 generated on the DynOmics server were selected using Principal Component Analysis (PCA). These structures were then used as structural bases for the construction and validation of molecular docking protocols in the Glide program. Molecular anchoring demonstrated that the conformation 3 results showed the highest rate of enrichment during screening of a set of active compounds and decoys. Then, the docking poses for POPTc80 inhibitors described in the literature were considered *queries* for generating and validating shape-based models in the vROCS program. As a result, a 3D structure of the POPTc80 was possible, a validated docking protocol was developed and a specific shape-based model for POPTc80 was also developed. At the end of this process, the best molecular docking protocols and the best shape-based model were used as computational filters for the virtual screening of the ChemBridge library, which made it possible to obtain a set of putative hits that will be experimentally validated in the POPTc80 enzyme in collaboration.

Keywords: Chagas disease, CADD, Homology modeling, Molecular docking, Virtual Screening.

1. INTRODUÇÃO

1.1. Doença de Chagas

A doença de Chagas, ou tripanossomíase americana, é uma doença tropical negligenciada causada pelo protozoário flagelado *Trypanosoma cruzi*, pertencente à ordem Kinetoplastida, família Trypanosomatidae (CHAGAS, 1909). A doença pode ser transmitida aos seres humanos de diversas formas. A via de transmissão vetorial por triatomíneos hematófagos conhecidos popularmente como barbeiros é a mais comum. Nesse caso, a contaminação ocorre através do contato das fezes (contaminadas com *T. cruzi*), produzidas pelo triatomíneo durante o repasto sanguíneo (a saliva do triatomíneo libera uma substância urticante no local da picada), com a lesão provocada pela picada ou mucosas (KIRCHHOFF, 2011). Transmissão por via oral, que ocorre quando há ingestão de alimentos (garapa, açaí, etc.) contaminados por triatomíneo ou suas fezes. Outras vias de transmissão possíveis, embora menos comuns, são transmissão vertical, transmissão transfusional e transplante de órgãos, além da transmissão por acidentes laboratoriais por material biológico contaminado (BERN, 2011; DNDI, 2021).

A infecção por *T. cruzi* apresenta duas fases clínicas: aguda e crônica. A fase aguda é caracterizada pela visualização do parasito no exame direto de sangue em microscópio, durante os três primeiros meses após a infecção. Esta fase geralmente é assintomática (~90% dos casos) ou oligossintomática (fase caracterizada por febre, fadiga, dores no corpo e edema facial) e, portanto, as triagens virtuais são valorizadas apenas nas áreas endêmicas. Como a doença não é diagnosticada precocemente, a fase aguda evolui para a fase crônica, a qual é caracterizada pela presença de anticorpos específicos (IgG anti-*T. cruzi*). Nesta fase, o parasito pode causar hiperplasia e graves danos teciduais aos órgãos do sistema circulatório ou trato gastrointestinal, resultando em um quadro clínico de insuficiência cardíaca e arritmia. Cerca de 30% dos indivíduos infectados cronicamente desenvolvem alterações cardíacas e até 10% deles desenvolvem alterações digestivas, neurológicas ou mistas que podem requerer tratamento específico. Em função das graves manifestações clínicas, muitos indivíduos infectados vão a óbito (LESCURE et al., 2010; PÉREZ-MOLINA et al., 2015).

1.1.1. Etiologia e distribuição

De acordo com a Organização Mundial de Saúde (OMS), a doença de Chagas acomete cerca de 6 milhões de pessoas no mundo, 173.000 novos casos por ano, 9.490 mortes em 2019, 75 milhões de pessoas sob risco, menos de 10% dos infectados foram diagnosticados, endêmico em 21 países na América Latina e também presente na América do Norte, Europa, Japão e Austrália, devido a imigração de pessoas de áreas endêmicas (Figura 1) (DNDI, 2021; LIDANI et al., 2019). Diante do carácter impactante e debilitante supracitado, uma parcela significativa dos indivíduos infectados perde a sua capacidade produtiva, causando fortes impactos sociais e econômicos. A identificação e tratamento de pacientes nos estágios iniciais resulta em um custo-benefício cumulativo de US\$ 7.273 a US\$224.981 no custo atual de identificação e tratamento (BARTSCH et al., 2018).

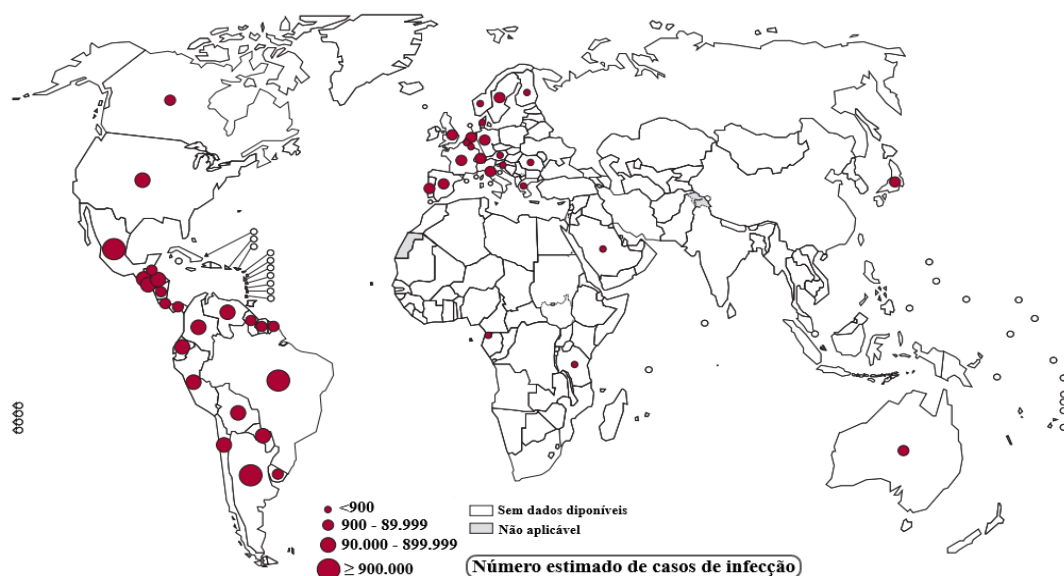


Figura 1- Distribuição global da doença de Chagas, baseada em estimativas oficiais, no período de 2006-2015 (Adaptado de WORLD HEALTH ORGANIZATION, 2020).

1.1.2. Ciclo Biológico

O *T. cruzi* apresenta três formas durante o seu ciclo biológico: amastigota, epimastigota e tripomastigota (Figura 3) (ORTEGA-BARRÍA, 2008).

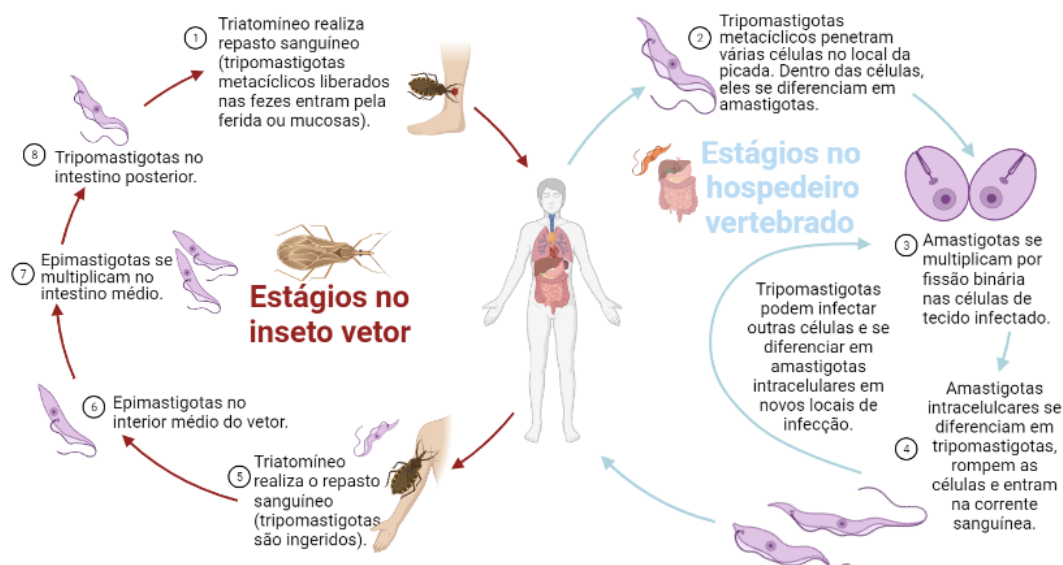


Figura 2 - Ciclo biológico de *Trypanosoma cruzi*. (1) O inseto triatomíneo realiza repasto sanguíneo e libera tripomastigotas metacíclicas nas fezes, os quais adentram o hospedeiro vertebrado através de abertura no local da picada ou mucosas intactas; (2) dentro do hospedeiro vertebrado, os tripomastigotas invadem células e se diferenciam em amastigotas intracelulares; (3) As amastigotas se multiplicam dentro das células; (4) Em seguida, as amastigotas intracelulares se diferenciam em tripomastigotas, rompem as células e são liberadas na corrente sanguínea, podendo adentrar novas células, se diferenciar novamente em amastigotas, e se replicarem; (5) O triatomíneo, ao realizar repasto sanguíneo, ingere tripomastigotas circulantes; (6) As tripomastigotas ingeridos se diferenciam em epimastigotas no intestino médio do vetor; (7) as epimastigotas se multiplicam no intestino médio do vetor; (8) Ao final deste processo, as epimastigotas se diferenciam em tripomastigotas metacíclicas no intestino posterior (Adaptado de (CDC, 2021).

A Figura 2 apresenta o ciclo biológico de *T. cruzi*. A contaminação do hospedeiro vertebrado geralmente através da contaminação de mucosas ou lesão gerada durante o repasto sanguíneo por triatomíneos hematófagos por fezes contendo as formas tripomastigotas metacíclicas infectantes. Dentro do hospedeiro vertebrado, os parasitos infectam diferentes tipos celulares, após se diferenciarem em amastigotas, se multiplicam no ambiente intracelular (KIRCHHOFF, 2011; ROBERTS; JAVONY, 2009). A forma amastigota (Figura 3-a) é a forma replicativa intracelular encontrada nos tecidos do hospedeiro vertebrado. Ela apresenta formato oval ou esférico e um flagelo pequeno, localizado dentro de uma bolsa (ORTEGA-BARRÍA, 2008). Após se

multiplicarem dentro das células hospedeiras, as amastigotas se diferenciam em tripomastigotas e, em seguida, rompem as células. A forma tripomastigota (Figura 3-c) é encontrada no sangue e apresenta formato alongado. Ela apresenta aspecto fusiforme, com o cinetoplasto localizado na sua porção anterior, próximo do núcleo. Além disso apresenta um flagelo na forma uma pequena membrana ondulante (BOGITSH; CARTER; OELTMANN, 2013; CASTRO, 2019). As tripomastigotas liberadas invadem tecidos adjacentes e se espalham hematogênicamente para diversos locais, nos quais eles iniciam novos ciclos assíncronos de multiplicação intracelular, mantendo assim a parasitemia em níveis suficientes para infectar novos triatomíneos (4–8 semanas durante a fase aguda) (FRANCO-PAREDES, 2014; PÉREZ-MOLINA; MOLINA, 2018). O triatomíneo se infecta ao ingerir sangue de mamíferos que apresentam tripomastigotas circulantes, que são formas infectantes e não replicativas. No intestino médio do vetor, os parasitos se diferenciam em epimastigotas, os quais se multiplicam no ambiente extracelular. A forma epimastigota (Figura 3-b) é a forma replicativa encontrada em triatomíneos hematófagos e não infectante para hospedeiros vertebrados (humanos). Após migrarem para o intestino posterior, as epimastigotas se diferenciam em tripomastigotas metacíclicas (formas não replicativas e infectantes), os quais são liberados junto às fezes do vetor durante o próximo repasto sanguíneo (KIRCHHOFF, 2011).

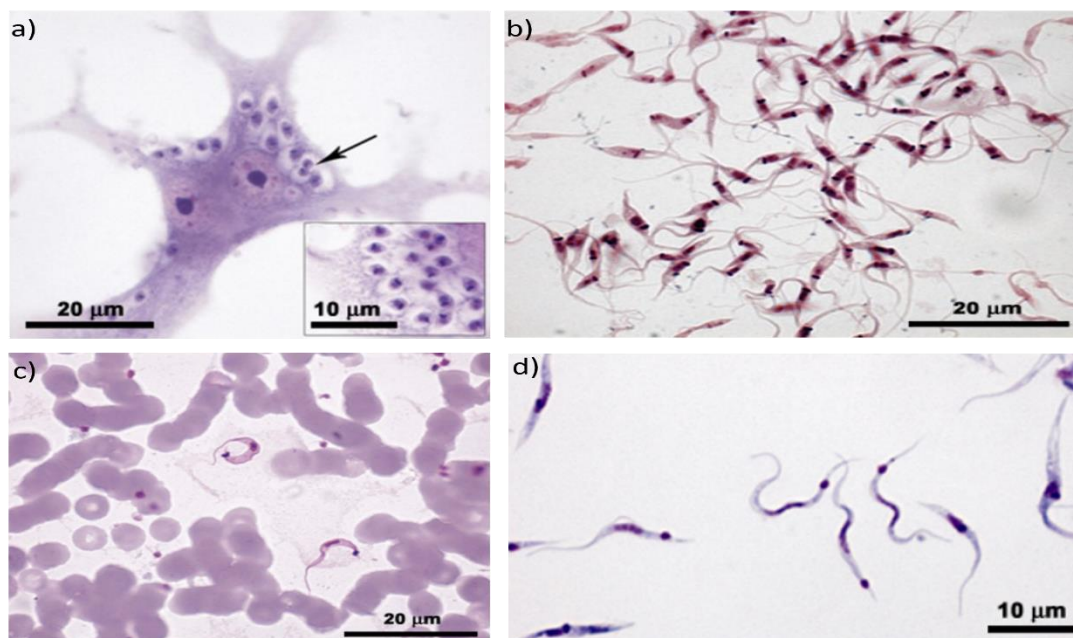


Figura 3 - Formas morfológicas do *T. cruzi*. Forma amastigota (a), epimastigota (b), tripomastigota circulante (c) e tripomastigota metacíclica (d) (FIOCRUZ, 2017).

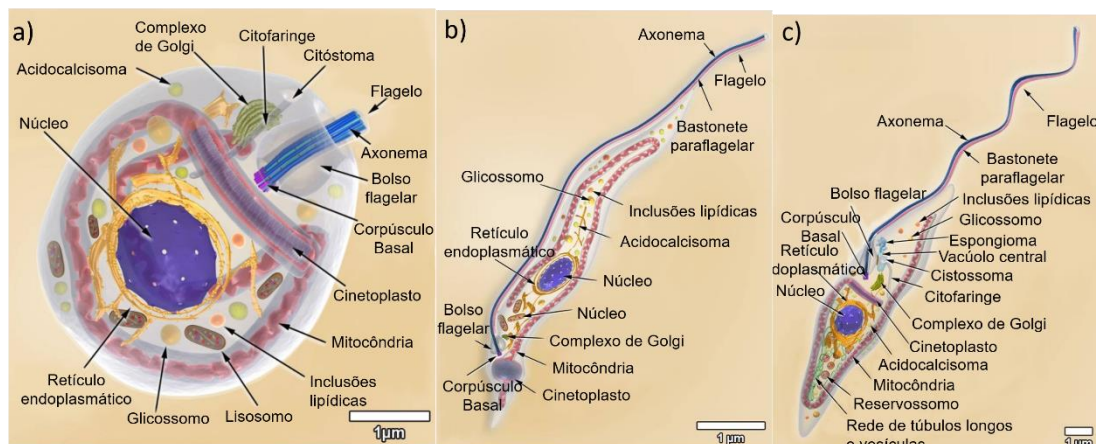


Figura 4 – Organelas celulares das formas Amastigota (a), epimastigota (b) e tripomastigota (c) (TEIXEIRA et al., 2012).

1.1.3. Tratamento farmacológico

Atualmente, apenas o fármaco benznidazol (BZN, 1) está disponível para o tratamento da doença de Chagas no Brasil. O BZN (Figura 4) é um composto nitroaromático absorvido por via oral, sendo uma absorção rápida e praticamente total, sendo excretado através das fezes e urina. Quando empregado adequadamente durante a fase aguda da doença, o BZN apresenta elevada eficácia, com taxas de cura em torno de 76% (CANÇADO, 1999). Por outro lado, o fármaco apresenta baixa eficácia na fase crônica, sendo de 60-93% em crianças até 13 anos (PORCEL et al., 1998; SGAMBATTI DE ANDRADE et al., 1996; YUN et al., 2009), e nos adultos, essa taxa varia de 2-40% (CANÇADO, 1999; FABBRO et al., 2007; VIOTTI et al., 2006).

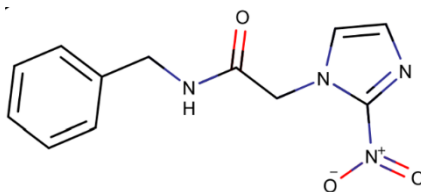


Figura 5 - Estrutura química do fármaco benznidazol (1), utilizado no tratamento da doença de Chagas.

O mecanismo de ação do BZN é dependente de metabolização enzimática (redução do grupo nitro) por nitroreduzases tripanossômicas tipo I (NTR), uma classe de enzimas insensíveis ao oxigênio presentes em vários protozoários parasitas, para os quais não há homólogo mamífero (KEENAN; CHAPLIN, 2015; PATTERSON; WYLLIE, 2014). Depois de várias reações, o BZN é reduzido a uma hidroxilamina, que é convertida através de várias transformações não enzimáticas em dialdeído glioxal (HALL; WILKINSON, 2012; WILKINSON; KELLY, 2009). Este altamente

reativo metabólito se liga a proteínas, lipídios, DNA/RNA e pequenas moléculas, como glutationa, bloqueando a síntese de novas fitas de DNA e inibindo sistema antioxidante do *T. cruzi*, tornando o parasito altamente suscetível ao dano oxidativo (CALDAS; SANTOS; NOVAES, 2019; MÜLLER KRATZ et al., 2018; PATTERSON; WYLLIE, 2014). O metabolismo do BZN também produz derivados do ânion do radical nitro, que causam efeitos citotóxicos contra o parasito (CALDAS; SANTOS; NOVAES, 2019; RAJÃO et al., 2014). No entanto, outros mecanismos potenciais têm sido sugeridas uma vez que o BZN é capaz de inibir proteínas, DNA e síntese de RNA em *T. cruzi* (HALL; WILKINSON, 2012; WILKINSON; KELLY, 2009). O fármaco também possui propriedades imunomoduladoras, induzindo a atividade fagocítica de macrófagos e células dendríticas, o que pode aumentar a depuração do parasito (CALDAS; SANTOS; NOVAES, 2019; MAYA et al., 2007).

Nas últimas décadas, preocupações associadas ao perfil toxicológico de BZN têm sido levantadas diante de seu potencial carcinogênico, teratogênico, genotóxico e toxicidade reprodutiva (DE TORANZO; MASANA; CASTRO, 1984; FERREIRA; FERREIRA, 1986; GORLA; CASTRO, 1985; TEIXEIRA; CALIXTO; TEIXEIRA, 1994). Esses resultados estão de acordo com os achados de outros fármacos nitroimidazólicos usadas na clínica, como o metronidazol. Esses efeitos tóxicos podem ser parcialmente impulsionados por espécies reativas de oxigênio decorrentes da redução do grupo nitro a um grupo amino por NTRs tipo II (sensíveis ao oxigênio), que estão presentes em organismos mamíferos (PATTERSON; WYLLIE, 2014; TROCHINE et al., 2014). Diante do exposto, considera-se que é extremamente necessário desenvolver novos fármacos mais eficazes, seguros, com e adequados para sistemas de saúde com recursos financeiros limitados (FIELD et al., 2017).

1.2. Alvos moleculares para o planejamento de novos fármacos

Diante dos avanços obtidos no campo da genômica, transcriptômica e proteômica (ATWOOD et al., 2005), diversas vias metabólicas de *T. cruzi* até então desconhecidas puderam ser estudadas como potenciais alvos biológicos para o planejamento de fármacos tripanocidas ou tripanostáticos (Figura 6).

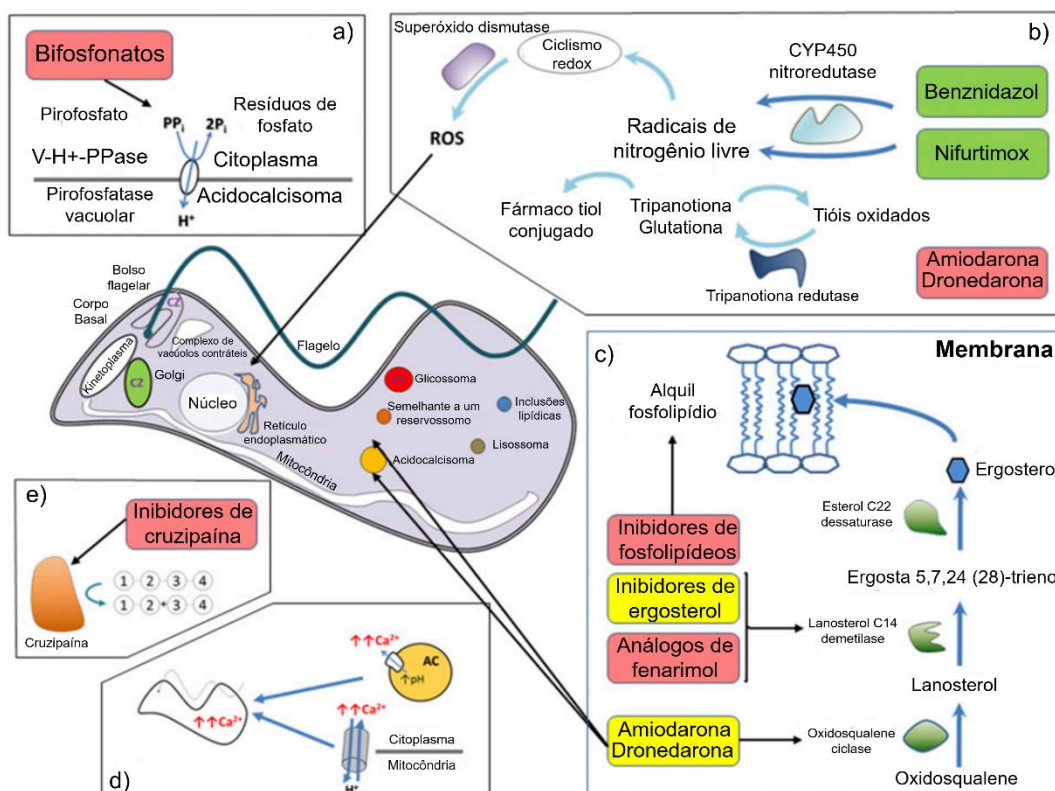


Figura 6 – Representação esquemática do modo de ação dos principais fármacos com atividade tripanocida. As caixas verdes representam fármacos atualmente usados para tratamento (exceto Nifurtimox), as caixas amarelas representam medicamentos em ensaios clínicos e as caixas vermelhas representam medicamentos experimentais. (a) Os bifosfonatos inibem a farnesil pirofosfato sintase, que reduz os níveis de esteróis e outros compostos poliisoprenoides essenciais, afetando a viabilidade das células. (b) Nifurtimox (NFX) e benznidazol (BZN) são reduzidos pela nitroredutase do parasito, resultando na produção de espécies de oxigênio reativo (ROS), que danificam diretamente as células do parasito. A tripanotiona redutase ajuda a aliviar o estresse oxidativo, e os inibidores dessas enzimas, como a tioridazina e a sulfoximina butionina, aumentam a quantidade de ROS no espaço intracelular. (c) O ergosterol biossintético é essencial para a sobrevivência do parasito. O bloqueio dessa via leva à perda da viabilidade celular por meio da depleção de esteróis essenciais e acúmulo de

intermediários tóxicos. Inibidores de ergosterol e análogos de fenarimol têm como alvo a lanosterol C14 demetilase, e amiodarona e dronedarona parcialmente inibe a oxidosqualeno ciclase. Os inibidores de fosfolipídios bloqueiam a síntese de esteróis, inibem a síntese de novo de fosfolipídios pela via de Greenberg e inibem enzimas de transdução de sinal como a fosfatidilinositol fosfolipase C. (d) Amiodarona e dronedarona liberam Ca^{2+} das mitocôndrias e acidocalcisomas, o que aumenta os níveis de Ca^{2+} no espaço citoplasmático e compromete a sobrevivência celular. (e) Cruzipaína (CZ) é tipicamente localizada no complexo de Golgi, bolsa flagelar e glicosomas e é uma cisteíno protease essencial envolvida na diferenciação do parasito, invasão celular, multiplicação e evasão imune. Os inibidores da cruzipaína alteram o complexo de Golgi, devido ao acúmulo de precursores da cruzipaína (SALES JUNIOR et al., 2017).

As proteases são uma classe de proteínas capazes de quebrar ligações peptídicas entre os aminoácidos de macromoléculas (BARRETT; RAWLINGS; O'BRIEN, 2001). As proteases são potenciais alvos farmacológicos devido as suas funções utilizada por protozoários que inclui: invasão de células hospedeiras e tecidos, degradação de mediadores da resposta imune e na hidrólise de proteínas do hospedeiro (KLEMBBA; GOLDBERG, 2002; MCKERROW et al., 1993; ROSENTHAL, 1999).

Várias proteases presentes em tripanossomatídeos (*Trypanosoma cruzi* e *Trypanosoma brucei*) e *leishmania spp* foram purificadas e caracterizadas bioquimicamente. Essas peptidases pertencem a 5 classes diferentes de proteases: cisteíno-proteases, serino-proteases, treonino-proteases, aspártico-proteases e metalo-proteases, sendo cada protease identificada no mínimo em um estágio de seu ciclo biológico em pelo menos uma espécie de tripanossomatídeo (ALVAREZ et al., 2021). Ao todo, mais de 40 famílias de proteases de tripanossomatídeos já foram descritas, abrindo um leque de possíveis alvos farmacológicos (ALVAREZ et al., 2021).

As cisteíno-proteases são antigamente conhecidas como tiol proteases, devido a presença do grupo tiol que atua como nucleófilo por estar localizado próximo a histidina do sítio ativo, que atua como próton (SAJID; MCKERROW, 2002; WILKESMAN, 2017). Essa classe requer a presença de uma cisteína essencial, tendo sua tríade formada por cisteína-histidina-aspartato (MURI, 2014), desempenhando funções no metabolismo, diferenciação, evasão da resposta imune e invasão de células (CAZZULO; STOKA; TURK V, 1997; FRANKE DE CAZZULO et al., 1994;

MEIRELLES et al., 1992). As principais cisteíno-proteases encontradas em *T. cruzi* são: cruzipaina e catepsina B-like, sendo a cruzipaina um alvo já estudado para o tratamento da doença de Chagas (SIQUEIRA-NETO et al., 2018).

As treonino-proteases são caracterizadas pela presença de uma treonina em sua tríade catalítica. O proteassoma é o complexo enzimático mais estudado nos tripanossomatídeos (ALVAREZ et al., 2021). A estrutura e propriedades bioquímicas são bem caracterizadas desde seu primeiro estudo (GONZÁLEZ et al., 1996; HUA et al., 1996). O proteassoma é fundamental na degradação de proteínas ubiquitina-dependente, cujo papel é desempenhado por várias subunidades com atividade de treonino-proteases (GOLDBERG; ROCK, 1992). Essas proteases já mostraram sua participação na diferenciação celular do parasito (CAZZULO, 2002; GONZÁLEZ et al., 1996), tornando essa classe susceptível à quimioterapia.

As aspártico-proteases são caracterizadas pela presença de um par de resíduos de ácido aspártico e essa classe possui preferência em clivar ligações peptídicas entre aminoácidos hidrofóbicos (DAVIES, 1990; RAWLINGS; BARRETT, 1995). Um aspártico-protease encontrada em *T. cruzi* é a DDi1-like, que é formada por três domínios: domínio N-terminal semelhante a ubiquitina, domínio semelhante a protease retroviral e um domínio C-terminal associado a ubiquitina (ALVAREZ et al., 2021). Estudos mostraram que a inibição da DDi1-like de *T. cruzi* é capaz de inibir o crescimento de epimastigotas.

As metalo-proteases são amplamente encontradas em procariotos e eucariotos, possuem um cátion bivalente, normalmente cálcio ou zinco, encontrado no sítio ativo que pode direcionar a ligação peptídica a ser clivada, estabilizando a própria proteína. Essas proteases foram descritas em *T. cruzi* pela primeira vez em 1996, onde foi detectado a expressão de um complexo arranjo de metalo-proteases, aparentemente ligado a membrana (LOWNDES et al., 1996). Logo após, vários estudos surgiram para tentar identificar e agrupar essas proteases (GRANDGENETT et al., 2000; MA et al., 2011). Várias metalo-proteases como a GP63, leucil aminopeptidases e entre outras já são alvos validados para tratamento de doenças como a doença de Chagas, Leishmanioses e outras (ALVAREZ et al., 2021).

Dentro da família das proteases, as serino-proteases desempenham papéis críticos em muitos processos fisiológicos e patológicos e são alvos diagnósticos e terapêuticos comprovados em uma série de indicações clínicas. A supressão das atividades proteolíticas aberrantes dessas proteases tem sido usada clinicamente para

o tratamento de doenças relevantes. Os polipeptídeos com 10-20 resíduos são de grande interesse como moduladores medicinais de serina proteases, porque esses peptídeos demonstram as características de fármacos de moléculas pequenas e fármacos macromoleculares (XU; HUANG, 2020).

Um grupo de serina peptidases, a família prolil oligopeptidase, não pode hidrolisar peptídeos contendo mais de cerca de 30 resíduos. Este grupo não está relacionado com as famílias clássicas de tripsina e subtilisina, e inclui dipeptidil peptidase IV, acilaminoacil peptidase e oligopeptidase B, além do protótipo prolil oligopeptidase (POLGÁR, 2002).

Prolil oligopeptidases foram investigados como potenciais alvos terapêuticos, bem como agentes terapêuticos. Desde a atividade dessas enzimas, especialmente no soro, foi correlacionado a distúrbios neurológicos, como depressão, mania, esquizofrenia, ansiedade, anorexia e bulimia nervosa (MAES, 1998; MAES et al., 1995, 2001), esforços estão sendo feitos para a busca ativa por inibidores específicos de prolil oligopeptidase.

1.2.1. Prolil oligopeptidase 80 de *T. cruzi*

Entre as proteases de *T. cruzi*, a prolil oligopeptidase 80 (POPTc80) destaca-se como um alvo promissor. A POPTc80 é uma serino-protease da família S9 (EC 3.4.21.26), ou seja, um grupo de enzimas que cliva a extremidade carboxila da prolina (BASTOS et al., 2005; GRELLIER et al., 2001). Esta oligopeptidase de 80 kDa secretada por tripomastigotas infectantes está envolvida na degradação específica de colágenos humanos purificados de tipo I e IV e a hidrólise nativa de colágeno tipo I no mesentério de ratos a pH fisiológico (MARTINS SANTANA et al., 1997).

Mecanicamente, o sucesso da infecção por *T. cruzi* depende da sua capacidade de degradar a matriz extracelular (ME) para ter acesso à célula hospedeira. Para isso, as tripomastigotas possuem na sua superfície proteínas (membros da família Tc-85) capazes de se ligarem a fibronectina/laminina, que são elementos da ME (GIORDANO et al., 1994, 1999). Em termos da degradação da ME, a POPTc80 já é consolidada para fazê-la (BASTOS et al., 2010), pois é secretada por formas tripomastigotas e hidrolisa colágenos dos tipos I e IV e fibronectina em vários sítios pós prolina (BASTOS et al., 2005; MARTINS SANTANA et al., 1997). Esta atividade colagenolítica foi demonstrada *in situ* sobre o mesentério de rato (tecido rico em

colágeno do tipo I) sendo comparável a colagenase de *Clostridium histolyticum*, bactéria altamente invasiva (MARTINS SANTANA et al., 1997).

Nas últimas décadas, diversos inibidores seletivos de POPTc80 (Figura 6) foram investigados *in vitro* para avaliar o possível papel da enzima na infecção de células hospedeiras não fagocíticas (BAL et al., 2003; JOYEAU et al., 2000; VENDEVILLE et al., 1999a). Essas moléculas inibem, *in vitro*, a entrada de tripomastigotas em diferentes tipos de células não fagocíticas, de forma dose dependente (BASTOS et al., 2005; GRELLIER et al., 2001), demonstrando que a POPTc80 pode estar envolvida na degradação de colágenos da matriz extracelular da célula infectada (DE ALMEIDA et al., 2016).

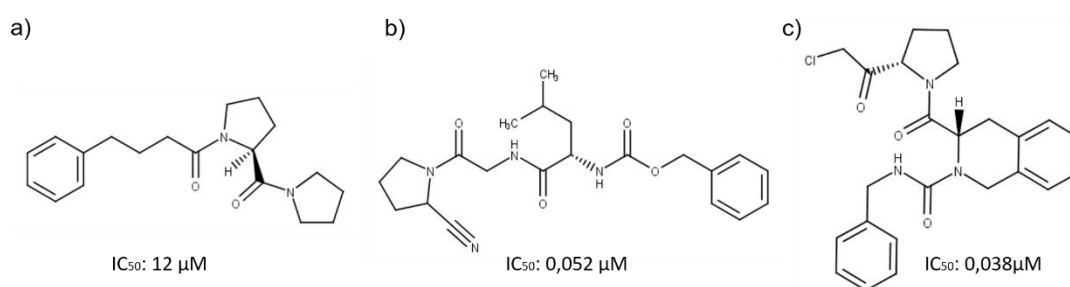


Figura 7 – Compostos com atividade inibitória para POPTc80 (BAL et al., 2003; JOYEAU et al., 2000; VENDEVILLE et al., 1999a).

Além disso, a POPTc80 é um alvo validado pelo nocaute de genes (*knockout*). Ambos os alelos dessa proteína (Non-Esmeraldo Like POPTc80 e Esmeraldo-Like POPTc80) foram nocauteados tanto por metodologia clássica quanto usando o sistema CRISPR/Cas9 (SILVA, 2018). A deleção dos dois alelos da POPTc80 indica que esse gene é essencial à viabilidade do parasito, devido à não sobrevivência da cultura em que foi realizado o duplo nocaute. Já a deleção de apenas um alelo, o parasito apresenta um crescimento lento e a sua diferenciação é afetada (SILVA, 2018). Esses achados reforçaram a importância da POPTc80 como um potencial alvo terapêutico para o planejamento de novos fármacos tripanostáticos.

Para complementar ainda mais a POPTc80 como um alvo promissor, esta enzima já foi encontrada em três diferentes cepas: Tulahuen, Y e Berenice (GRELLIER et al., 2001). Outro estudo recente demonstrou que as vacinas à base de POPTc80 são capazes de conferir proteção contra a infecção por *T. cruzi*, e esta imunoproteção foi extensiva à diferentes cepas (CL Brener, Tulahuen, SylvioX10/1 e Dm28c) pertencentes a diferentes DTUs (do inglês, *discrete typing units*) (BIVONA et al., 2018). Diante desses fatos, a POPTc80 demonstra diversos sinais sobre sua

importância no ciclo biológico do parasito e sua importância para o sucesso de infecção celular.

A POPTc80, assim como outras POPs apresentam dois domínios (Figura 8): domínio β -propeller composto de sete pás (folhas β coloridas em azul claro); e domínio α/β hidrolase formado por hélices α e folhas β (representado na coloração salmão). O domínio catalítico é a interface entre os dois domínios (α/β hidrolase e β -propeller). Formado pela porção N- e C-terminal, tem o arranjo clássico de α/β hidrolase, enquanto o domínio β -propeller as folhas β estão dobradas e arranjadas radialmente em volta de um túnel central, no entanto, este túnel é estabilizado por ligações hidrofóbicas entre a primeira e última pá (FÜLÖP; SZELTNER; POLGÁR, 2000).

A tríade catalítica (Ser548, Asp631, His667) está localizada em uma grande cavidade na interface entre os dois domínios (FÜLÖP; BÖCSKEI; POLGÁR, 1998), e sua função catalítica, mostrado na Figura 6, apresenta o buraco de oxinônio formado pelas serinas 548 e 546. A serina ataca o plano trigonal carbono carbonílico cindível (Figura 7a), formando um oxinônio intermediário tetraédrico (Figura 7b). O intermediário tetraédrico colapsa, o que é seguido pela perda do peptídeo C-terminal. Isso gera uma enzima acil intermediário (Figura 7c). O intermediário acil é então atacado por uma molécula de água ativada, dando origem ao segundo intermediário tetraédrico (Figura 7d). Este intermediário então colapsa levando a liberação do peptídeo N-terminal (EKICI; PAETZEL; DALBEY, 2008). A Serina é encontrada na ponta de uma curva muito acentuada, referida como cotovelo nucleófilo, que é característico das hidrolases do tipo α/β (OLLIS et al., 1992). Consequentemente, o grupamento hidroxila da serina está bem exposto e facilmente acessível ao grupo do imidazol catalítico de um lado e ao substrato do outro. A Histidina é encontrada no meio de um loop. Um dos átomos de oxigênio do Aspartato está no plano do anel imidazol da Histidina, proporcionando a posição ideal para a formação da ligação de hidrogênio, e também é ligada por hidrogênio a uma molécula de água bem ordenada (POLGÁR, 2002).

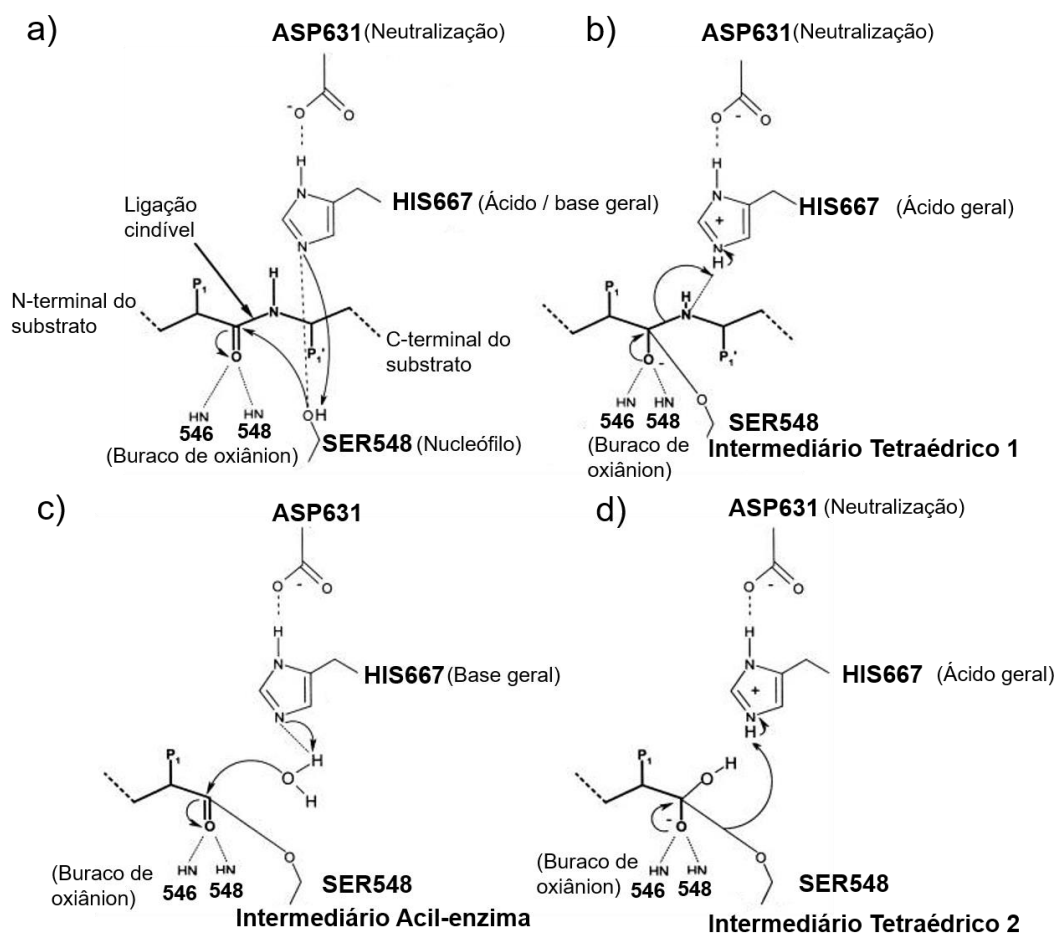


Figura 8 - O mecanismo proteolítico utilizado pelas serina proteases “clássicas” Ser/His/Asp. As etapas do mecanismo catalítico de protease de serina Ser/His/Asp "clássico" incluem: complexo de Michaelis (a), intermediário tetraédrico 1 (b), intermediário acil-enzima (c) e intermediário tetraédrico 2 (d)(EKICI; PAETZEL; DALBEY, 2008).

A POPTc80 possui dois tipos de conformação: fechada e aberta (Figura 8). A conformação de qualquer proteína é importante para a sua função biológica. As proteínas possuem uma conformação ativa, que é a conformação de menor energia (ou mínimo global), correspondendo a conformação em que a proteína desempenha a sua função biológica (VERLI, 2014). Para uma proteína interagir com um ligante, é importante: i) alto nível de complementariedade estérica, ou seja, a proteína e o ligante possuem uma alta porcentagem de contatos moleculares, definidas pelos raios de van der Waals atômicos, em contato próximo; ii) alta complementariedade de propriedades associadas às superfícies de contato moleculares (esta complementariedade pode ser tanto eletrostática, onde grupos polares/carregados do ligante ficam perto de grupos da proteína com polaridade/carga complementar, quanto relacionada à

complementaridade de regiões hidrofóbicas); iii) o ligante geralmente se liga em uma conformação energeticamente favorável, e iv) interações repulsivas entre ligante e proteínas são minimizadas (VERLI, 2014).

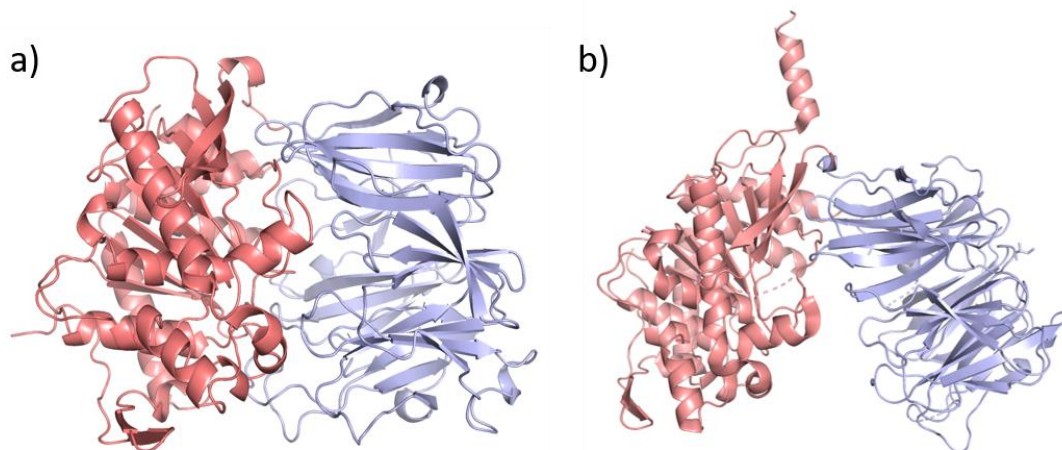


Figura 9 - Modelos estruturais de Prolil Oligopeptidase. Modelo estrutural da POPTc80 (a) gerado pelo AlphaFold, a POPTc80 é dividida em um domínio α/β hidrolase catalítico (salmão) composto por α -hélices e folhas β , e um domínio β -propeller (azul claro) formado somente por formado somente por folhas β . Estrutura da POP de *Sphingomonas capsulata* aberta (b)(SHAN; MATHEWS; KHOSLA, 2005).

1.3. Planejamento e descoberta de novos fármacos

A pesquisa e desenvolvimento (P&D) de fármacos constitui um processo complexo, competitivo, que demanda tempo, elevado investimento financeiro e integração de diversas áreas estratégicas como inovação, tecnologia, gerenciamento, etc (DIMASI; GRABOWSKI; HANSEN, 2015; LOMBARDINO; LOWE, 2004). Todo este processo leva em média 12 anos e normalmente custa 2,6 bilhões de dólares (CHAN et al., 2019). A figura 9 apresenta as principais etapas envolvidas no processo de P&D de fármacos.

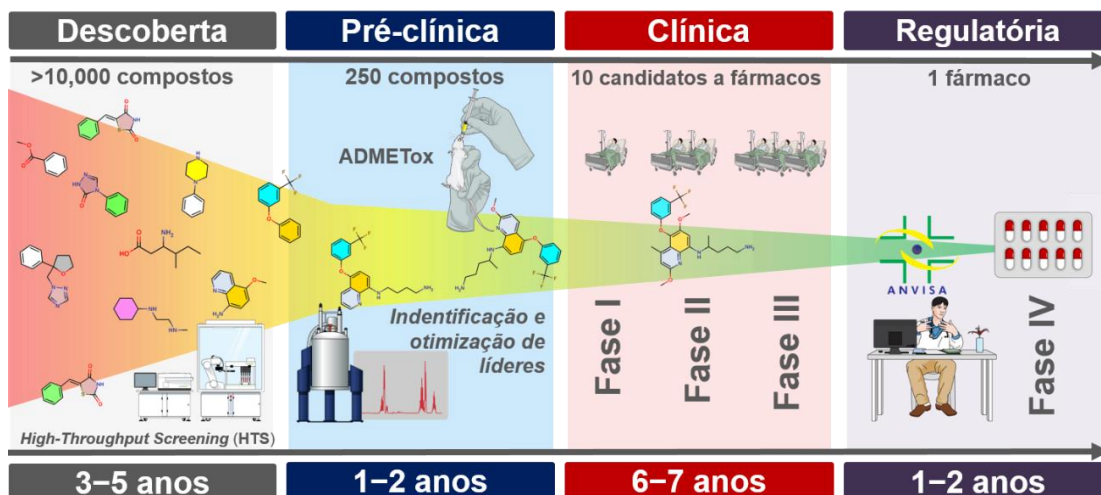


Figura 10 – Etapas envolvidas na pesquisa e desenvolvimento de fármacos (Cortesia do Prof. Dr. Bruno Junior Neves).

O processo de descoberta de fármacos pode ser dividida em três grandes etapas: descoberta, pré-clínica e clínica (ELEBRING; GILL; PLOWRIGHT, 2012; LOMBARDINO; LOWE, 2004).

A etapa de descoberta se inicia com a identificação de um potencial alvo biológico relacionado a determinado quadro clínico ou doença e, no caso de doenças parasitárias, relacionado à sobrevivência do parasito e estabelecimento do mesmo no organismo do hospedeiro. Em geral, o alvo biológico consiste em determinada proteína (macromolécula), por exemplo, uma enzima, receptor ou transportador. Preferencialmente, este alvo deve ser validado, ou seja, deve demonstrar um papel importante no processo fisiopatológico estudado (GASHAW et al., 2011; LOMBARDINO; LOWE, 2004; NWAKA; RIDLEY, 2003). Diferentes técnicas podem ser usadas na validação de alvos biológicos, como técnicas que bloqueiam a expressão de genes que codificam o alvo (gene *knock out*, por exemplo), observando a influência desses genes no processo fisiopatológico (HUGHES et al., 2011). Em doenças infecciosas e parasitárias, a validação do alvo consiste em se avaliar o impacto, para o agente etiológico, da não expressão de determinados genes que codificam determinado alvo, evidenciando assim a importância do mesmo para a sobrevivência do parasito ou microrganismo (CROWTHER et al., 2010; NWAKA; RIDLEY, 2003).

Após eleição e validação do alvo biológico, milhares de moléculas são experimentalmente testadas utilizando ensaios *in vitro* com alta vazão de dados (HTS, do inglês, *High Throughput Screening*) ou triagem virtual (VS, do inglês, *Virtual*

Screening) com o objetivo de identificar um subgrupo de compostos (~250) com atividade biológica *in vitro* (*hits*). Duas abordagens amplas são usadas para a descoberta de *hits*: (i) abordagens baseadas em alvos, ou seja, a identificação de inibidores de um alvo bioquímico específico; ou (ii) abordagens fenotípicas, as quais envolvem a triagem de inibidores do crescimento do parasito estudado, geralmente em culturas *in vitro*.

Durante a etapa pré-clínica, compostos líderes são estruturalmente planejados e otimizados a partir dos *hits* para ter propriedades físico-químicas, farmacocinéticas, farmacodinâmicas e toxicológicas adequadas. Posteriormente, estes protótipos são selecionados como novas entidades químicas para prosseguir para a fase de desenvolvimento clínico em seres humanos (HUGHES et al., 2011).

Os estudos clínicos são os testes feitos em seres humanos e podem ser divididos em três fases. Na fase I, os candidatos a fármacos, que foram bem sucedidos nos estudos pré-clínicos, são testados em um pequeno grupo de voluntários sadios (20-80 indivíduos). Essa fase tem como objetivo o estabelecimento de informações sobre a dosagem segura e informações a respeito do perfil farmacocinético (ADMET) dos compostos. Os estudos de fase II são realizados em um grupo maior de indivíduos portadores da doença ou quadro clínico investigado (200-300 indivíduos). Nesta etapa, os estudos são realizados para obter informações sobre segurança e o perfil inicial de efeitos adversos. Além disso, resultados preliminares de eficácia são analisados. Na fase III, os testes são feitos em grande escala, testando o composto em mais de 1000 pacientes (1000-3000 pacientes). Nesse momento, a eficácia do candidato a fármaco é estabelecida e os efeitos adversos menos comuns são observados. Assim que as informações necessárias a respeito da potência e eficácia do composto são elucidadas, os resultados são submetidos a um órgão regulador para que o novo fármaco seja aprovado e comercializado (LOMBARDINO; LOWE, 2004; NICOLAOU, 2014).

Ultimamente, estima-se que sejam necessários cerca de 8,3 anos para a introdução de um único fármaco no mercado e um investimento orçamentário que pode chegar a US\$ 985 milhões (WOUTERS; MCKEE; LUYTEN, 2020). O processo de P&D também apresenta elevada taxa de insucesso, uma vez que apenas 20% dos candidatos a fármacos em estudos de Fase III chegam ao mercado (HAY et al., 2014). A competitividade e o potencial de investimento em P&D também estão diretamente relacionados com a descoberta de *blockbusters* (termo em inglês utilizado para designar fármacos inovadores com vendas anuais superiores a US\$ 1 bilhão por ano)

(DEBNATH; AL-MAWSAWI; NEAMATI, 2010; JA, 2013). Consequentemente, as doenças tropicais negligenciadas tradicionalmente não integram este quadro de investimento uma vez estando associadas a condições de pobreza e sendo endêmicas apenas em países subdesenvolvidos.

1.3.1. Planejamento de fármacos auxiliado por computador

A utilização de ferramentas computacionais para a descoberta, planejamento e otimização novos protótipos de fármacos consiste em um campo de estudo em constante expansão denominado planejamento de fármacos auxiliado por computador (CADD, do inglês *Computer-Assisted Drug Design*) (WERMUTH et al., 1998). Estas ferramentas apresentam impacto significativo no processo de P&D, ao passo que os projetos podem ser desenvolvidos com infraestrutura, tempo e investimentos relativamente reduzidos. As entidades químicas com propriedades físico-químicas, farmacocinéticas e toxicológicas inadequadas também podem ser identificadas ou estruturalmente otimizadas ainda nas etapas iniciais do estudo, evitando assim fracassos durante os estudos clínicos.

A expansão do uso do CADD se deve, principalmente, aos grandes avanços, obtidos nas últimas décadas, nas áreas de *software* e *hardware*. Aliado a esses avanços, também se destaca o crescimento observado nas áreas de genômica e proteômica, que permitiu o acesso a um grande número de informações estruturais de alvos biológicos (BAJORATH, 2012; KAPETANOVIC, 2008; TALELE; KHEDKAR; RIGBY, 2010). Consequentemente, os últimos 15 anos presenciaram um aumento significativo no número de estruturas de proteínas depositadas no PDB (do inglês, Protein Data Bank). Atualmente, essa base de dados possui mais de 155 mil estruturas 3D experimentalmente determinadas por cristalografia de raio-X, por ressonância magnética nuclear (NMR) e/ou microscopia (GOODSELL et al., 2020). Em paralelo, graças aos avanços em química combinatória e HTS, houve também um aumento considerável no volume de dados de bioensaios. Atualmente, milhões de compostos com dados de bioatividade estão depositados em bases de dados de domínio público como PubChem Bioassay (KIM et al., 2019) e ChEMBL (MENDEZ et al., 2019). Consequentemente, esses dados podem ser utilizados como bases moleculares para construção e validação de modelos computacionais (BRAGA et al., 2014).

1.3.2. Predição de estruturas de proteínas

Graças aos avanços em biologia estrutural, o crescimento de dados estruturais de alvos biológicos vem crescendo através da evolução dos métodos de predição das estruturas 3D (SCHWEDE, 2013). Apesar disso, a cobertura do espaço biológico-estrutural ainda é bastante pequena frente aos milhões de proteínas. Atualmente, o próprio *T. cruzi* possui apenas 268 estruturas 3D disponíveis no PDB (WATERHOUSE et al., 2018). Consequentemente, métodos computacionais vêm sendo amplamente utilizados para prever as estruturas 3D de potenciais alvos terapêuticos de fármacos. Conforme descrito na Figura 10, as estruturas 3D de proteínas podem ser preditas com base em três tipos de abordagens:

- (i) Modelagem por homologia ou modelagem comparativa: abordagem baseada em molde (*template*) em que todos os membros de uma família de proteínas (identidade sequencial >25%) possuem o mesmo padrão estrutural (WATERHOUSE et al., 2018);
- (ii) *Threading* ou predição por enovelamento: abordagem baseada em molde (*template*) que na busca por sequências de aminoácidos (identidade sequencial >20%) consenso com elucidação experimental que apresentam um padrão estrutural (YANG; ZHANG, 2015);
- (iii) *Ab initio*: abordagem totalmente independente de *template* (estruturas homólogas), sem utilização de perfis ou modelos para a sua predição (VERLI, 2014).

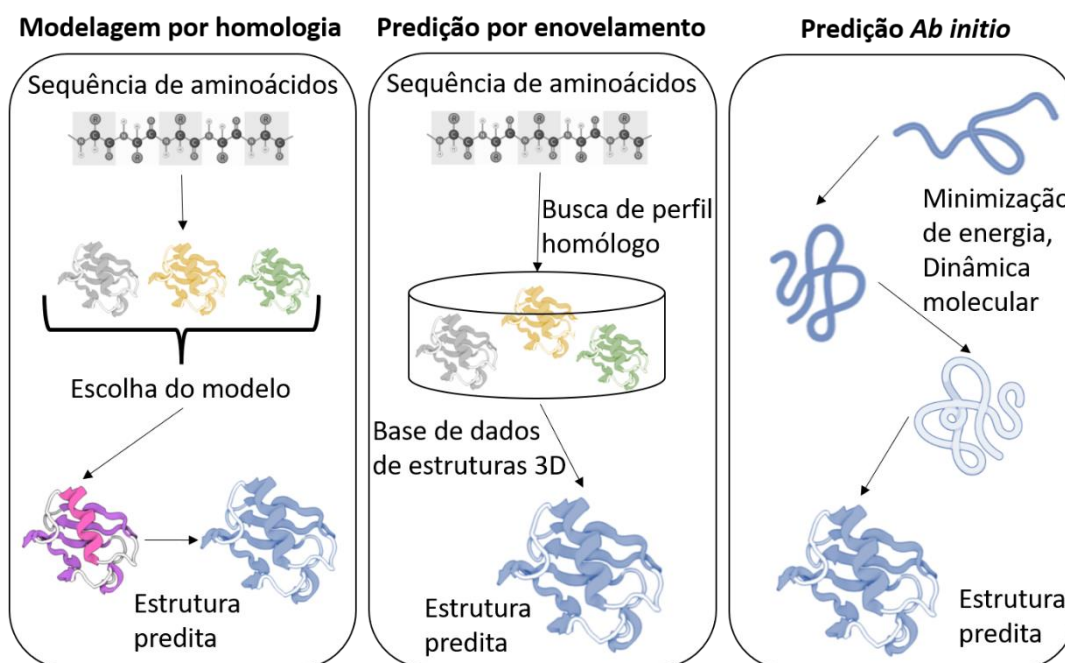


Figura 11 – Abordagens para predição de estrutura 3D. Modelagem por homologia, Predição por enovelamento e Predição *Ab initio*.

Atualmente, os servidores SWISS-MODEL (WATERHOUSE et al., 2018), I-TASSER (YANG; ZHANG, 2015) e AlphaFold (JUMPER et al., 2021) estão entre as ferramentas mais populares de modelagem por homologia, *threading* e *ab initio*, respectivamente, para predição de estruturas 3D de proteínas.

O servidor SWISS-MODEL (WATERHOUSE et al., 2018) gera um modelo 3D de uma sequência alvo através da informação estrutural extrapolada de uma estrutura de proteína relacionada à evolução, servindo como modelo. A predição do SWISS-MODEL consiste em cinco passos principais: dado de entrada, busca de modelo, escolha de modelo, construção de modelo e qualidade estimada do modelo. Na etapa do dado de entrada, a proteína alvo pode ser disponibilizada como sequência de aminoácidos (FASTA ou código UniProtKB). Na busca de modelo, o dado fornecido na primeira etapa serve como *query* para a busca de estruturas de proteínas relacionadas a biblioteca de modelos do SWISS-MODEL, e isso é realizado utilizando o método de alinhamento BLAST e o HHblits, que adiciona sensibilidade no caso de homologia remota. Na etapa de seleção do modelo, os modelos são classificados de acordo com a qualidade esperada dos modelos resultantes, conforme estimado pelo Modelo Global de Qualidade Estimada (GMQE) e Estimativa de Qualidade da Estrutura Quaternária (QSQE). Os modelos e alinhamentos mais bem classificados são comparados para verificar se eles representam estados conformacionais alternativos

ou cobrem diferentes regiões da proteína alvo. Na etapa de construção do modelo, um modelo de proteína 3D é gerado automaticamente pelas coordenadas do átomo conservado do primeiro anel de transferência, conforme definido pelo alinhamento do modelo alvo. Coordenadas de resíduos correspondentes a inserções/deleções no alinhamento são geradas por modelagem de alça e um modelo de proteína de átomo completo é obtido pela construção das cadeias laterais de aminoácidos não conservados. Para quantificar erros de modelagem e fornecer estimativas sobre a precisão esperada do modelo, o SWISS-MODEL conta com a função de pontuação QMEAN na etapa de qualidade estimada do modelo. QMEAN usa potenciais estatísticos de força média para gerar estimativas de qualidade globais e por resíduo. As estimativas de qualidade local são aprimoradas pela distância entre pares que representam informações de conjunto de todas as estruturas de modelo encontradas.

O servidor I-TASSER (YANG; ZHANG, 2015) é baseado no I-TASSER, um método hierárquico baseado em modelo para a estrutura de proteínas e previsões de função consistindo em três etapas gerais. Para uma dada sequência de consulta, I-TASSER primeiro identifica modelos estruturais ou estruturas secundárias da biblioteca PDB usando LOMETS, um programa de meta-*threading* que consiste em algoritmos de *threading* múltiplos. A topologia dos modelos de comprimento total é então construída remontando as estruturas de fragmento alinhadas continuamente excisadas dos modelos, onde as estruturas das regiões não alinhadas são construídas do zero por desdobramento *ab initio* com base em simulações de Monte Carlo de troca de réplicas. As trajetórias da estrutura são agrupadas por SPICKER para identificar estados de baixa energia livre. A partir do SPICKER clusters, uma segunda rodada de remontagem da estrutura é conduzida para refinar os modelos estruturais. As conformações de baixa energia livre são posteriormente refinadas por simulações atômicas completas usando FG-MD e ModRefiner. Finalmente, os insights funcionais da proteína de consulta são obtidos combinando o modelo estrutural com proteínas na biblioteca de funções BioLiP por meio de comparação de estrutura e perfil de sequência (YANG; ZHANG, 2015).

A ferramenta AlphaFold (JUMPER et al., 2021) melhora muito a precisão da predição estrutural ao incorporar uma abordagem baseada em inteligência artificial (redes neurais profundas ou *Deep Learning*) que promete desempenho superior na precisão de predição. A rede AlphaFold prevê diretamente as coordenadas 3D de todos

os átomos pesados para uma determinada proteína usando a sequência de aminoácidos primária e sequências alinhadas de homólogos como entradas (JUMPER et al., 2021).

A rede compreende duas etapas principais. Primeiro, o tronco da rede processa as entradas por meio de camadas repetidas de um novo bloco de rede neural que chamamos de Evoformer para produzir uma matriz $N_{seq} \times N_{res}$ (N_{seq} , número de sequências; N_{res} , número de resíduos) que representa pares de resíduos. A representação MAS é inicializada como o MAS bruto. As principais inovações no bloco Evoformer são novos mecanismos de troca de informações dentro do MAS e representações de pares que permitem raciocínio direto sobre as relações espaciais e evolutivas. O tronco da rede é seguido pelo módulo de estrutura que introduz uma estrutura 3D explícita na forma de rotação e tradução para cada resíduo da proteína. Essas representações são inicializadas em um estado trivial com todas as rotações definidas para a identidade e todas as posições definidas para a origem, mas rapidamente se desenvolvem e refinam uma estrutura altamente precisa com detalhes atômicos precisos. As principais inovações nesta seção da rede incluem quebrar a estrutura da cadeia para permitir o refinamento local simultâneo de todas as partes da estrutura, um novo transformador equivariante para permitir que a rede raciocine implicitamente sobre os átomos da cadeia lateral não representados e um termo de perda que coloca um peso substancial na correção da orientação dos resíduos. O refinamento iterativo usando toda a rede (que é chamado de ‘reciclagem’ e está relacionado a abordagens em visão computacional) contribui significativamente para a precisão com pouco tempo extra de treinamento (JUMPER et al., 2021).

1.3.3. Ancoramento molecular

O ancoramento molecular é uma abordagem de SBDD que tem como objetivo prever a posição, orientação (também chamada de pose) e afinidade (energias de pontuação) de ligação de ligantes no sítio ativo de alvos macromoleculares (ŠLEDŽ; CAFLISCH, 2018; TORRES et al., 2019).

As principais etapas envolvidas no ancoramento molecular estão representadas na Figura 11. O processo de ancoramento se inicia com a escolha de um alvo que tenha sua estrutura 3D determinada, seja por métodos experimentais como cristalografia de raio-X ou ressonância magnética nuclear (RMN), ou por métodos computacionais como modelagem por homologia (LIMA et al., 2016). Em seguida, uma caixa 3D conhecida popularmente como *grid* é inserida nas coordenadas do sítio ativo para

delimitar o espaço de investigação de complementaridade molecular (YOUNG, 2009). Na etapa seguinte, um algoritmo de busca exploratória explora diferentes orientações e conformações (poses) de um ou mais ligantes, separadamente, dentro do *grid*. Em paralelo, uma função de pontuação (*score*) estima a afinidade de cada uma das poses pelo *grid* (macromolécula).

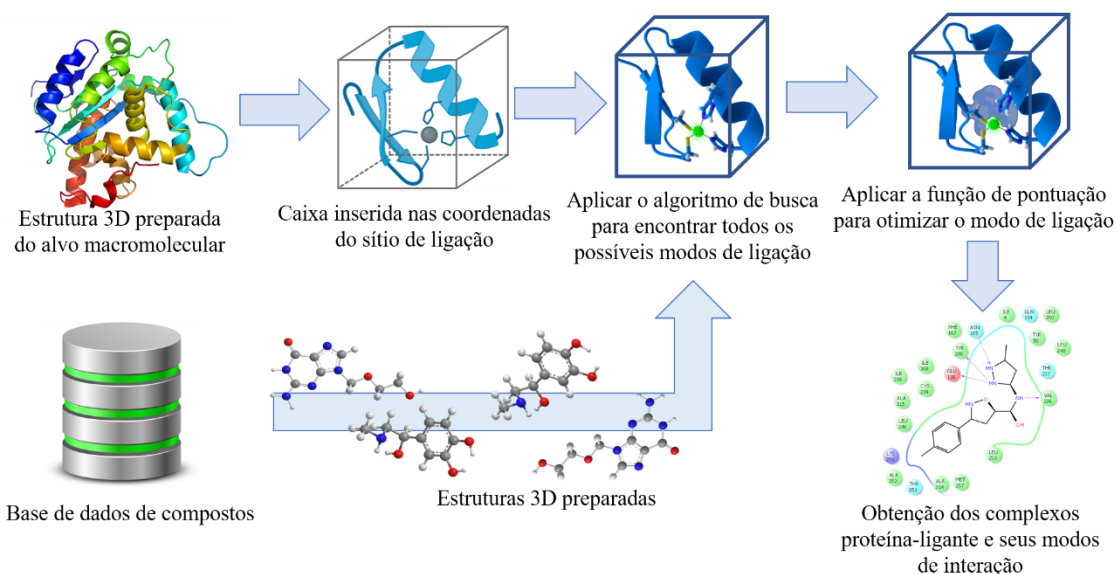


Figura 12 – Fluxograma de um típico ancoramento molecular.

Os algoritmos de busca exploratória podem ser classificados em três grupos (KITCHEN et al., 2004; VERLI, 2014).

- (1) **Métodos sistemáticos:** estes métodos exploram todos os graus de liberdade do ligante por meio de construção incremental, busca conformacional, ou por bancos de dados. A construção incremental pode ser feita de várias maneiras, acoplando fragmentos da região do sítio e crescendo-os através de ligações covalentes ou dividindo os fragmentos em uma parte rígida (fragmento de núcleo) e uma parte flexível (cadeia lateral) e inserindo um núcleo rígido no sítio de ligação e gradualmente adicionando cadeias laterais (DESJARLAIS et al., 1986; YOUNG, 2009). A busca conformacional é feita para encontrar o ligante com o confôrmero com menor energia global, que corresponde à conformação nativa (GUEDES; DE MAGALHÃES; DARDENNE, 2014). Outro método de pesquisa sistemática é o uso de uma biblioteca de conformações pré-geradas (LEACH; KUNTZ, 1992; YOUNG, 2009);
- (2) **Métodos estocásticos:** estes métodos modificam randomicamente o grau de liberdade do ligante (translacional, rotacional e conformacional) e, a cada

passo, geram uma diversidade de soluções. Cada pose do ligante recém-obtida é avaliada com base em uma função de probabilidade predefinida que verifica se a nova pose é melhor que a anterior e se irá mantê-la ou descartá-la (GOODSELL et al., 1993; HART; READ, 1992). Os métodos populares são: Algoritmos Genéticos (GA-*Genetic Algorithms*), Evolução Diferencial (DE-*Differential Evolution*), Algoritmos Meméticos (MA-*Memetic Algorithms*), Optimização por Enxame de Partículas (OEP), Arrefecimento Simulado (SA-*Simulating annealing*), Algoritmo de Colônia de Formigas, entre outros (KITCHEN et al., 2004);

- (3) **Métodos determinísticos:** esta abordagem utiliza métodos de minimização de energia e dinâmica molecular. Essas metodologias unem-se de ferramentas como o gradiente (direção de maior crescimento de uma superfície, por exemplo) e a movimentação do sistema ao longo do tempo, respectivamente. Portanto, garantem sempre convergência para um mínimo de energia. Nesses métodos, o estado atual do sistema determina as alterações a serem realizadas guiando para o próximo estado (GUEDES; DE MAGALHÃES; DARDENNE, 2014). Contudo, esses métodos exigem um alto custo computacional e muitas vezes leva a seleção indesejada das conformações para energia mínima local (BROOIJMANS; KUNTZ, 2003).

Por outro lado, as funções de pontuação podem ser classificadas em quatro grandes grupos (RASHID et al., 2013);

- (1) **Funções baseadas em campo de força:** são derivadas de campo de força clássico, que consiste no somatório dos termos de energia exponencial (átomos ligados e não ligados) de um sistema (complexo ligante-macromolécula). Esse tipo de função considera a energia de interação do complexo receptor-ligante, e a energia interna do ligante (dos termos ligados e não ligados). O principal problema dessa classe é a forte repulsão atômica em pequenas distâncias, derivada da aproximação clássica intrínseca ao campo de força (KITCHEN et al., 2004);
- (2) **Funções de pontuação empíricas:** são baseadas na energia de interação e nas informações estruturais, os coeficientes destes vários termos podem ser obtidos por meio de análise de regressão. Por conseguinte, através da soma dos termos individuais é possível realizar a aproximação da energia de ligação (ELDRIDGE et al., 1997);

- (3) **Funções baseadas em conhecimento (*knowledge-based*):** As funções de pontuação baseadas em conhecimento visam reproduzir a estrutura experimental em vez de energias de ligação. Nessas funções os complexos proteína-ligantes são modelados através da energia potencial de pares de ligações atômicos. Essas interações são definidas de acordo com um ambiente molecular em que os átomos se encontram dependendo do seu tipo atômico (MUEGGE, 2000). Portanto, como os métodos empíricos, as funções de pontuação baseadas em conhecimento tentam capturar implicitamente os efeitos de ligação que são difíceis de modelar explicitamente (YOUNG, 2009);
- (4) **Funções baseadas em aprendizado de máquina:** Desde 2010, várias funções de pontuação baseadas em aprendizado de máquina (*Machine learning*) foram introduzidas com sucesso em estudos de ancoramento molecular para aprender a partir da dependência não linear do complexo proteína-ligante e, conseqüentemente, dar uma previsão mais precisa da afinidade de ligação (KHAMIS; GOMAA; AHMED, 2015; SHEN et al., 2020). O aprendizado de máquina é visto como um subcampo da inteligência artificial que usa técnicas estatísticas para construir modelos matemáticos de vários tipos de dados. Uma vez desenvolvido, o modelo de aprendizado de máquina pode ser usado para fazer previsões ou decisões sem ser explicitamente programado para isso (EKINS et al., 2019).

Desde sua primeira utilização na década de 1980 (KUNTZ et al., 1982), o ancoramento molecular se popularizou como a ferramenta de CADD mais utilizada em projetos de P&D de fármacos, seja na indústria ou na academia (MORRIS; LIM-WILBY, 2008). Esta ferramenta apresenta várias vantagens estratégicas, tais como: baixo custo computacional, velocidade de processamento, além da variedade de algoritmos utilizados no ancoramento molecular. Por outro lado, o ancoramento molecular apresenta as seguintes limitações: baixa acurácia das funções de pontuação em relação ao valores de afinidade experimental, efeito aditivo das funções de pontuação, flexibilidade conformações limitada apenas aos ligantes ou alguns aminoácidos no sítio ativo (GUIDO; OLIVA; ANDRICOPULO, 2008; MENG et al., 2011).

1.3.4. Modelos de forma e volume molecular

O modelo de forma e volume molecular (do inglês, *shape-based*) é uma técnica de LBDD baseada no princípio de que moléculas com forma e volume semelhantes podem apresentar propriedades biológicas semelhantes. Conforme descrito na Figura 12, o processo de geração desse tipo de modelo envolve o alinhamento de estruturas 3D de moléculas à uma *query* (composto bioativo conhecido), através de um processo de otimização de corpo sólido que maximiza o volume de sobreposição entre elas. Durante esta etapa, apenas átomos pesados de um ligante são considerados, ou seja, todos os hidrogênios são ignorados (SWANN et al., 2011). Em seguida, uma função gaussiana suave é utilizada para representar o volume molecular, de modo que seja possível minimizar rotineiramente para a melhor correspondência global (GRANT; GALLARDO; PICKUP, 1996; HAWKINS; SKILLMAN; NICHOLLS, 2007). Características como doadores e aceptores de ligação de hidrogênio, grupos ou átomos carregados positivamente (cátions) e negativamente (ânions), anéis aromáticos e grupos hidrofóbicos podem ser utilizadas para maximizar a sobreposição que leve a melhor correspondência global (SWANN et al., 2011).

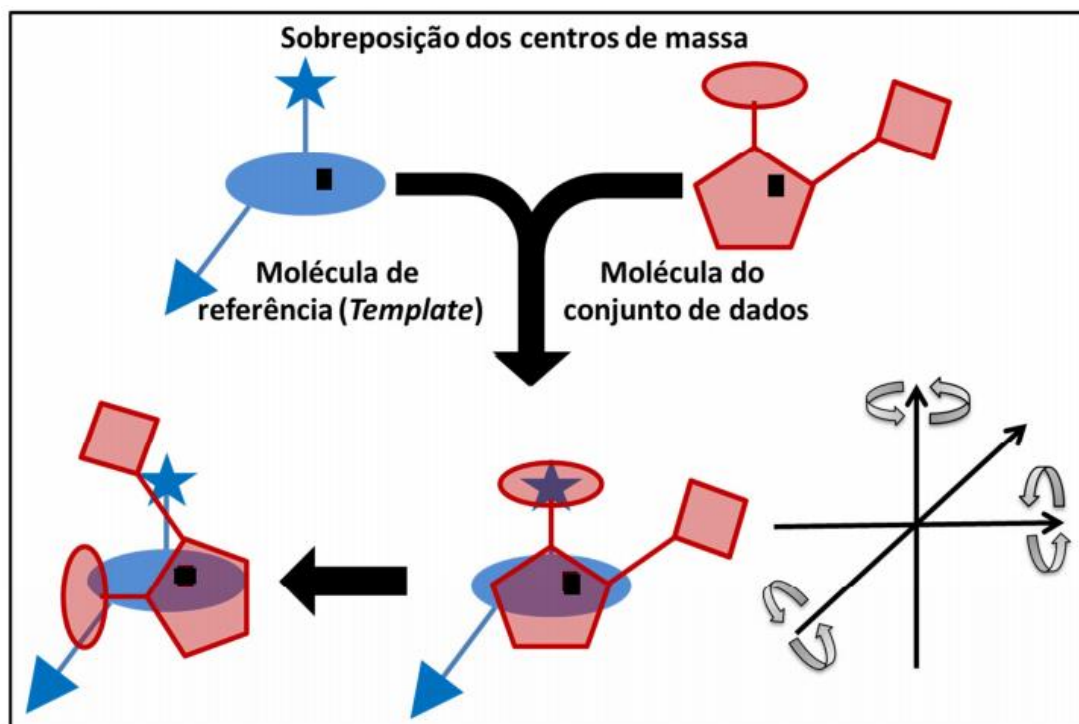


Figura 13 - Representação da geração de modelos baseados na forma e volume molecular no programa ROCS (OPENEYE SCIENTIFIC SOFTWARE, 2020).

1.3.5. Triagem Virtual

A triagem virtual (VS) (Figura 13) é uma estratégia de CADD poderosa para orientar a identificação de um subconjunto de *hits* virtuais ($10^1 - 10^2$ estruturas químicas) com potencial atividade biológica a partir de grandes bibliotecas virtuais de compostos ($10^3 - 10^9$ estruturas químicas) utilizando filtros computacionais (SCHNEIDER, 2010). Duas abordagens de VS são comumente empregadas: as baseadas em alvos (SBVS, do inglês *Structure-Based Virtual Screening*) e as baseadas em ligantes (LBVS, do inglês *Ligand Based Virtual Screening*). Na SBVS, métodos de SBDD como ancoramento molecular são utilizados como filtros computacionais para a priorização de *hits* virtuais, ao passo que na LBVS métodos de LBDD como modelos *shape-based* são utilizados para orientar a seleção de compostos (BADRINARAYAN; NARAHARI SASTRY, 2011; BRAGA; ANDRADE, 2013; RASHID et al., 2013; SCIOR et al., 2012; SHOICHET, 2004; WOLBER; LANGER, 2005).

Uma vez conduzida com o devido cuidado, as taxas de acerto de VS geralmente chegam a 40%, enquanto as taxas de recuperação de *hits* em triagens experimentais randômicas, tais como HTS, variam entre 0,01% e 0,1%. Apesar disso, a expectativa de que a abordagem de VS possa substituir completamente os ensaios experimentais *in vitro* e *in vivo* é excessivamente otimista. Portanto, a avaliação experimental *in vitro* dos *hits* virtuais priorizados deve ser considerada como a etapa final e mais importante de qualquer estudo VS (POSNER; XI; MILLS, 2009).

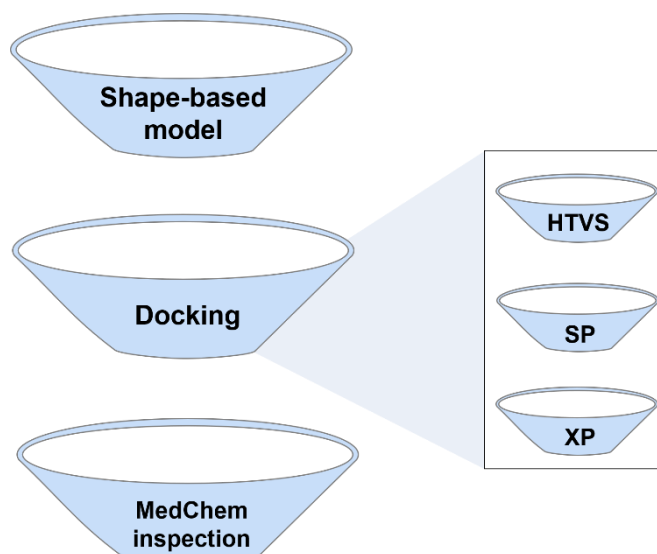


Figura 14 – Etapas envolvidas em uma típica triagem virtual.

2. JUSTIFICATIVA

Tendo em vista que essa doença foi descoberta há mais de um século e, apesar disso, o único fármaco aprovado no Brasil apresenta baixa eficácia e perfil toxicológico perigoso. Diante disso, a descoberta de novos fármacos com mecanismos de ação inovadores para tratamento da doença de Chagas é eminente. Graças aos avanços no conhecimento dos mecanismos envolvidos na relação parasito-hospedeiro, diversos alvos biológicos essenciais para manutenção do processo fisiopatológico e infecção foram identificados. Entre eles, a POPTc80 destaca-se como um alvo estratégico para rompimento da matriz extracelular, impedindo a disseminação e infecção das formas tripomastigotas. Diante do exposto, o planejamento e identificação de inibidores da POPTc80 explorado nesta dissertação justifica-se pelas seguintes razões:

- A POPTc80 é um alvo de *T. cruzi* validado geneticamente de pelo nocaute de gene (SILVA, 2018). Além disso, ensaios fenotípicos demonstraram que com inibidores seletivos de POPs a entrada de tripomastigotas em diferentes tipos de células não fagocíticas, de forma dose dependente (GRELLIER et al., 2001).
- A estrutura 3D de POPTc80 predita por métodos computacionais, tais como modelagem por homologia, *threading* e modelagem *ab initio*, pode servir como “base estrutural” para o planejamento e descoberta de inibidores de POPTc80. Nesse contexto, as propriedades eletrônicas, estéricas e volume molecular do sítio ativo da estrutura 3D POPTc80 podem ser utilizados para explorar a complementaridade ligante-macromolécula;
- As abordagens de CADD exploradas nesta dissertação representam ferramentas custo-efetivas para identificar novos inibidores de POPTc80 com investimentos de infra-estrutura, tempo e orçamento reduzido;
- A descoberta de novos inibidores de POPTc80 capazes de bloquear a infecção celular representa um bom ponto de partida para o *hit-to-lead* e otimização de compostos líderes com atividade tripanostática. Além disso, tal descoberta abre novas possibilidades para associação com o fármaco benznidazol.

3. OBJETIVOS

3.1. Objetivo geral

O objetivo geral desta dissertação é identificar novos inibidores da enzima POPTc80 através da integração métodos computacionais baseados na estrutura do alvo de do ligante.

3.2. Objetivos específicos

Os objetivos específicos dessa proposta são:

- a) Predizer a estrutura 3D da proteína POPTc80 utilizando modelagem por homologia, *threading* e modelagem *ab initio*;
- b) Construir e validar estatisticamente protocolos de ancoramento molecular a partir da estrutura 3D de POPTc80 e um conjunto de inibidores (ativos) e *decoys*;
- c) Investigar a influência de diferentes conformações de POPTc80 na taxa de enriquecimento dos protocolos ancoramento molecular, com vistas a determinar a conformação bioativa da enzima;
- d) Construir e validar modelos baseados em forma e volume molecular utilizando como *queries* um conjunto de inibidores de POPTc80 disponíveis na literatura científica;
- e) Realizar uma VS da biblioteca comerciais ChemBridge (1,3 milhões de compostos) utilizando os modelos computacionais gerados, a fim de selecionar um subconjunto de potenciais inibidores de POTc80 com atividade tripanostática;
- f) Realizar a avaliação experimental dos compostos priorizados através de ensaios enzimáticos.

4. MÉTODOS

4.1. Métodos computacionais

4.1.1. Predição da estrutura 3D

Considerando a ausência de estruturas POPTc80 resolvidas e disponibilizadas na base de dados *Protein Data Bank* (PDB), sua estrutura 3D foi predita computacionalmente. Inicialmente, a sequência FASTA de aminoácidos da POPTc80 (Código de identificação: TcCLB.506247.230) foi extraída da base de dados TriTrypDB (ASLETT et al., 2009) e submetida como arquivo de entrada (input) para predição estrutural nos servidores SWISS-MODEL (BORDOLI et al., 2009), I-TASSER (YANG; ZHANG, 2015) e AlphaFold (JUMPER et al., 2021).

4.1.2. Refinamento estrutural e validação da POPTc80

Os modelos 3D de POPTc80 preditas anteriormente foram então refinados estruturalmente utilizando o servidor GalaxyREFINE (KO et al., 2012). O GalaxyREFINE realiza uma otimização iterativa com vários operadores geométricos e em seguida promove o relaxamento da estrutura 3D utilizando dinâmica molecular para melhorar a qualidade geométrica estrutural (KO et al., 2012). Ao final deste processo, a qualidade dos ângulos phi (ϕ) e psi (ψ) de cadeia principal (análise de Ramachandran) e ângulos chi (χ) dos rotâmeros de cadeia lateral foi investigada utilizando o servidor MolProbity (CHEN; POHLHAUS, 2010; HINTZE et al., 2016).

4.1.3. Análise estrutural da POPTc80

Ao final da validação estrutural, a melhor estrutura de POPTc80 foi submetida ao processo de geração de conformações utilizando um modelo de rede anisotrópico (ANM, do inglês *Anisotropic Network Model*) (LI et al., 2017), disponível no servidor *DynOmics* (LI et al., 2017). A ANM é uma ferramenta de NMA (do inglês, *Normal Mode Analysis*) que permite a análise de movimentos vibracionais e a flexibilidade da proteica em sistemas moleculares. Ela utiliza uma rede elástica (EN, do inglês, *Elastic Network*) que representa a proteína como um grafo, no qual os carbonos alfas ($C\alpha$) dos aminoácidos são representados como nós conectados por molas e o potencial geral é simplesmente a soma dos potenciais harmônicos entre os nós em interação. A EN inclui todas as interações dentro de uma distância de corte. As informações sobre a orientação de cada interação em relação ao sistema de coordenadas global são consideradas dentro da matriz da constante de força e permitem a previsão de

movimentos anisotrópicos (LI et al., 2017). A geração de conformações foi realizada utilizando o módulo padrão (default) do servidor DynOmics, como a distância de corte (7,3Å), fator de escala constante da mola (1), expoente de escala de distância para constante de mola (0).

Em seguida, as 29 conformações obtidas a partir da ANM foram submetidas a análise de componentes principais (PCA, do inglês *Principal Component Analysis*) (RINGNÉR, 2008), disponível no pacote Bio3D (GRANT; SKJÆRVEN; YAO, 2021), implementado em linguagem de programação R v.4.1.2 (CHAN, 2018). A PCA transforma as variáveis originais (combinação linear das variáveis originais) em variáveis latentes com dimensionalidade reduzida, denominadas componentes principais (PCs, do inglês *Principal Components*). Este método é particularmente útil para destacar padrões conformacionais na estrutura 3D de POPTc80. Conseqüentemente, é possível revelar movimentos estruturais coletivos capturados pelos melhores PCs e determinar como estes estão frequentemente relacionados à função da proteína (GRANT; SKJÆRVEN; YAO, 2021).

4.1.4. Predição dos sítios de ligação na POPTc80

As conformações de POPTc80 selecionadas previamente foram então submetidas a predição dos sítios de ligação no servidor FTSite (KOZAKOV et al., 2015). Este servidor utiliza um conjunto de 16 sondas moleculares diferentes com tamanho e polaridade variados (Tabela 1) em uma grade densa ao redor da estrutura 3D da proteína. Através das sondas é possível identificar os potenciais bolsos de interação usando funções empíricas de energia livre. Em seguida, as sondas individuais (Tabela 1) são agrupadas e os agrupamentos são classificados com base na energia livre média. Os agrupamentos de consenso são classificados com base no número total de interações intermoleculares (KOZAKOV et al., 2015).

Tabela 1 - Conjunto de sondas moleculares utilizadas pelo FTSite.

Molécula-sonda	Características físico-químicas
Acetamida (ACD)	Polar, doadora e acceptora de ligação de hidrogênio
Acetonitrila (ACN)	Polar, acceptora de ligação de hidrogênio
Acetona (ACT)	Polar, acceptora de ligação de hidrogênio

Acetaldeído (ADY)	Polar, acceptora de ligação de hidrogênio
Metanamina (AMN)	Polar, doadora e acceptora de ligação de hidrogênio
Benzaldeído (BDY)	Polar, aromática, acceptora de ligação de hidrogênio
Benzeno (BEN)	Hidrofóbica, aromática
tert-butanol (BUT)	Hidrofóbica, aromática
Ciclohexano (CHX)	Polar, doadora e acceptora de ligação de hidrogênio
<i>N,N</i> -dimetilformamida (DFO)	Polar, acceptora de ligação de hidrogênio
Dimetil éter (DME)	Polar, acceptora de ligação de hidrogênio
Etanol (EOL)	Polar, doadora e acceptora de ligação de hidrogênio
Etano (ETH)	Hidrofóbica
Fenol (PHN)	Polar, aromática, doadora e acceptora de ligação de hidrogênio
Isopropanol (THS)	Polar, acceptora de ligação de hidrogênio
Ureia (URE)	Polar, positiva doadora e acceptora de ligação de hidrogênio

4.1.5. Avaliação do nível de conservação dos aminoácidos da POPTc80

O servidor ConSurf (ASHKENAZY et al., 2016) foi utilizado para realizar a análise de avaliação do nível de conservação dos aminoácidos da POPTc80. O servidor ConSurf estima a conservação evolutiva de posições de aminoácidos em uma proteína com base nas relações filogenéticas entre sequências homólogas. Na presente dissertação, o grau de conservação dos aminoácidos da POPTc80 foi estimado usando 150 homólogos de outros organismos identificados na base de dados UNIREF-90 (SUZEK et al., 2015) (esta base de dados agrupa sequências e subfragmentos com 11 ou mais resíduos que pelo menos 90% de identidade de sequência entre si (de qualquer organismo) em uma única entrada UniRef, exibindo a sequência de um representante através do algoritmo de busca HMMER (ASHKENAZY et al., 2016). As sequências foram agrupadas e aquelas apresentando similaridade de sequência alta (>95%) foram excluídas usando o algoritmo CD-HIT para filtrar sequências redundantes (ASHKENAZY et al., 2016). Da mesma forma, as sequências com identidade

sequencial <35% ou >95% foram ignoradas. Um alinhamento de sequência múltipla (MSA) das sequências homólogas foi construído usando o método MAFFT-L-INS-I, e a árvore filogenética foi construída usando o algoritmo *Rate4Site* (PUPKO et al., 2002). As pontuações de conservação específicas da posição foram calculadas usando o método empírico Bayesiano (ASHKENAZY et al., 2016).

4.1.6. Compilamento e preparo do conjunto de validação

Um conjunto de 31 compostos com dados de atividade inibitória sobre a POPTc80 com compilado a partir da literatura científica (BAL et al., 2003; DE ALMEIDA et al., 2016; GRELLIER et al., 2001; VENDEVILLE et al., 1999a, 1999b). Em seguida, um limiar de atividade de 10 μM foi considerado para categorizar os compostos em ativos e inativos. Consequentemente, 24 compostos foram categorizados em ativos (IC_{50} ou $\text{K}_i \leq 10 \mu\text{M}$) e 7 como inativos (IC_{50} ou $\text{K}_i > 10 \mu\text{M}$). Considerando a proporção artificial entre ativos e inativos, 36 *decoys* para cada composto ativo foram introduzidos no conjunto de dados. *Decoys* são compostos assumidamente inativos, embora não tenham sido testados experimentalmente. Eles possuem propriedades físico-químicas semelhantes às dos compostos ativos, de forma que a taxa de enriquecimento na validação de modelos computacionais não seja apenas uma separação superotimística baseada em propriedades irrelevantes. Por outro lado, o *decoys* devem ser estruturalmente dissimilares dos compostos ativos, de modo que sejam provavelmente inativos (BRAGA et al., 2014; BRAGA; ANDRADE, 2013; NEVES et al., 2021).

Na presente dissertação, 864 *decoys* foram selecionados a partir do conjunto de dados DrugNow (16.647.845 compostos), disponível na base de dados ZINC15 (STERLING; IRWIN, 2015), utilizando fluxo de trabalho *in house* desenvolvido em *workspace* KNIME v. 4.3.3. (BERTHOLD et al., 2009). Os *decoys* são compostos assumidamente inativos, que deve se assemelhar as propriedades físicas dos ligantes ativos, porém, os *decoys* também devem possuir propriedades químicas distintas dos compostos ativos (HUANG; SHOICHET; IRWIN, 2006). A seleção de *decoys* obedeceram às seguintes regras de similaridade físico-química em relação ao espaço químico de 24 compostos ativos: massa molecular ($\pm 25 \text{ Da}$), número de ligações rotacionáveis (± 1), número de doadores de ligação de hidrogênio (± 1), número de aceptores de ligação de hidrogênio (± 2) e coeficiente de partição óleo-água ($\log P \pm 1$). Além disso, os *decoys* obrigatoriamente apresentaram coeficiente de Tanimoto (cT)

$<0,75$ (*fingerprints* ECFP4) com os compostos ativos e $cT <0,9$ entre os próprios *decoys*. Esse fluxo de seleção foi baseado no protocolo padronizado para geração de *decoys* disponível na DUD (do inglês, *directory of useful decoys*) (HUANG; SHOICHET; IRWIN, 2006).

4.1.7. Preparo do conjunto de validação

As estruturas químicas dos compostos ativos e *decoys* foram importadas para a plataforma Maestro v. 11.5 (Schrödinger, LLC, New York, NY, 2016) e preparadas utilizando o programa LigPrep (Schrödinger, LLC). O estado de protonação foi calculado de acordo com pH fisiológico ($7,4 \pm 0,5$) utilizando o programa Epik v. 4.3 (SCHRÖDINGER, 2009; SHELLEY et al., 2007), e a geometria molecular otimizada usando o campo de força OPLS3 (HARDER et al., 2016).

4.1.8. Preparo da proteína

As estruturas 3D da POPTc80 foram importadas para a plataforma Maestro v. 11.5 (Schrödinger, LLC) e preparadas utilizando o módulo *Protein Preparation Wizard* (MADHAVI SASTRY et al., 2013). Para o preparo da proteína, átomos de hidrogênio foram adicionados, a ordem de ligação e carga formais foram ajustadas. Os estados de protonação dos aminoácidos polares foram preditos em pH neutro (7.4 ± 0.5) utilizando o programa Epik v. 4.3 (SHELLEY et al., 2007) para representar as mesmas condições experimentais do ensaio enzimático. Além disso, a estrutura foi submetida a otimização das orientações de hidrogênio com pH 7.4, utilizando o programa PROPKA.

4.1.9. Ancoramento molecular

Estudos de ancoramento molecular foram realizados utilizando os módulos HTVS (do inglês, *High-throughput Virtual Screening*), SP (do inglês, *Standard Precision*) e XP (do inglês, *Extra Precision*) do programa Glide v. 7.8 (FRIESNER et al., 2004, 2006), disponível na plataforma Maestro v. 11.2 (Schrödinger, LLC). Previamente ao ancoramento, grids virtuais foram gerados para três conformações diferentes de POPTc80 geradas previamente e considerando as nas coordenadas x , y e z de dois potenciais sítios de ligação, conforme descrito na Tabela 2.

Tabela 2 – Coordenadas dos grids construídos para a POPTc80 em ambos os sítios.

Conformação da POPTc80	Coordenadas x , y e z	
	Sítio A	Sítio B

Conformação 09	21,83Å; 22,4Å; 21,60Å	27,67Å; 46,67Å; 43,00Å
Conformação 15	23,33Å; 23,33Å; 23,00Å	27,69Å; 46,67Å; 43,42Å
Conformação 20	22,26Å; 21,92Å; 21,65Å	27,97Å; 46,84Å; 43,05Å

O ancoramento molecular foi realizado considerando os ligantes flexíveis e a proteína na forma rígida. As poses de ancoramento foram ranqueadas utilizando três funções de pontuação: *Docking Score*, *Glide Energy* e *Glide Emodel*. O *Docking Score* (ou *GlideScore*) é uma função de pontuação empírica baseado no *ChemScore*, mas inclui um termo de conflito estérico, adiciona termos polares concebidos pela Schrödinger para penalizar as incompatibilidades eletrostáticas, e tem modificações em outros termos: O *Glide Emodel* combina o *score* de energia do *grid* e a afinidade de ligação predita pelo *GlideScore* (FRIESNER et al., 2004). Já o *Glide Energy* utiliza a energia de interação de Coulumb-van der Waals modificada (SCHRÖDINGER, 2009).

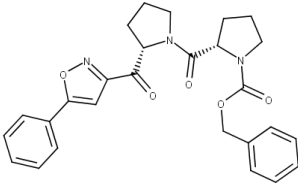
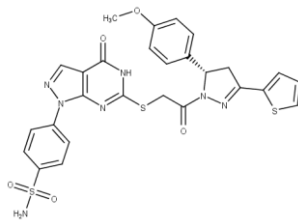
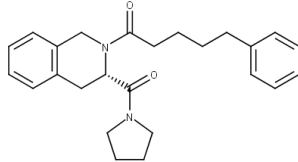
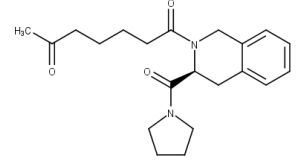
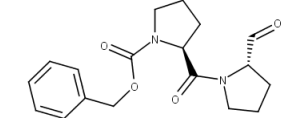
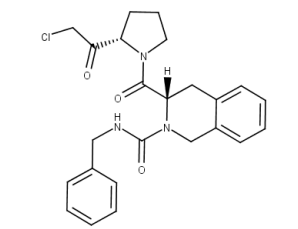
4.1.10. Modelos *shape-based*

Inicialmente, 500 conformações foram geradas para cada composto do conjunto de validação utilizando o programa OMEGA v.3.1.0.3 (Openeye Scientific Software, Santa Fe, NM, <http://www.eyesopen.com>) (HAWKINS et al., 2010). Em seguida, cargas atômicas parciais AM1-BCC (JAKALIAN; JACK; BAYLY, 2002) foram calculadas para cada conformação utilizando o programa QUACPAC v.2.0.0.3 (Openeye Scientific Software, Santa Fe, NM, <http://www.eyesopen.com>).

Em seguida, modelos *shape-based* foram desenvolvidos para predição de inibição de POPTc80 utilizando apenas as informações de forma e volume molecular. Nesta etapa, as poses de ancoramento de 6 compostos ativos (Tabela 3), assumidas como conformações potencialmente bioativas, foram carregadas no programa vROCS v. 3.4.2.1 (OpenEye Scientific Software, Santa Fe, NM, <http://www.eyesopen.com>) (SWANN et al., 2011) e utilizadas como *queries* (moldes) para definição das formas e volumes moleculares. Os parâmetros utilizados para a geração do modelo *shape-based*, foi a seleção do campo de força ExplicitMillsDean (MILLS; DEAN, 1996). Este campo de força leva em consideração os estados de protonação e carga do arquivo de entrada de inibidores e não-inibidores. Para o ranqueamento dos modelos compostos, foi utilizada a função *RefTverskyCombo* (KEARNES; PANDE, 2016), com a aleatoriedade, do inglês “*Random*”, de 50, permanecendo as configurações adicionais de fábrica do programa. A função *RefTverskyCombo* é uma combinação

das funções *RefTversky* (compara as moléculas de acordo com a melhor sobreposição do volume molecular) e *RefColorTversky* (relacionada à sobreposição apropriada de grupos de cores, representada que apresentam características como doadores e aceptores de ligação de hidrogênio, hidrofóbicos, cátions, ânions e anéis). O valor de *RefTverskyCombo* pode variar de 0 a 2, sendo que valores mais próximos a 2 indicam uma boa sobreposição das moléculas (MCGAUGHEY et al., 2007).

Tabela 3 – Compostos utilizados para a criação de modelos *Shape-based*.

Composto (query)	Estrutura 2D	Dado de atividade	Fonte
Query 1		Ki = 0,00021 μM	(BAL et al., 2003)
Query 2		Ki = 1,009 μM	(DE ALMEIDA et al., 2016)
Query 3		IC ₅₀ = 0,014 μM	(VENDEVILLE et al., 1999a)
Query 4		IC ₅₀ = 0,025 μM	(VENDEVILLE et al., 1999b)
Query 5		Ki = 0,00021 μM	(LAWANDI et al., 2010)
Query 6		IC ₅₀ = 0,0038 μM	(GRELLIER et al., 2001)

4.1.11. Validação estatística dos modelos gerados

A fim de avaliar a taxa de enriquecimento dos protocolos de ancoramento molecular e modelos *shape-based* gerados, as seguintes métricas foram calculadas: área sob a curva ROC (AUC), o fator de enriquecimento (EF) e a discriminação aprimorada de Boltzmann (BEDROC). Todos os parâmetros foram analisados em um workflow desenvolvido *in house* e implementado na plataforma KNIME versão 3.7.

A análise ROC (do inglês, *receiver operating characteristic*) é um método de avaliação de modelos amplamente utilizado, uma vez que fornece um resultado visual e numérico (Figura 14). A curva obtida nessa análise, conhecida como curva ROC, é construída a partir da variação de limiares (*thresholds*), apresentando no eixo y os valores de sensibilidade (taxa de verdadeiros positivos) e no eixo x as taxas de falsos positivos (1- especificidade), correspondentes a cada limiar (SONEGO; KOCSOR; PONGOR, 2008).

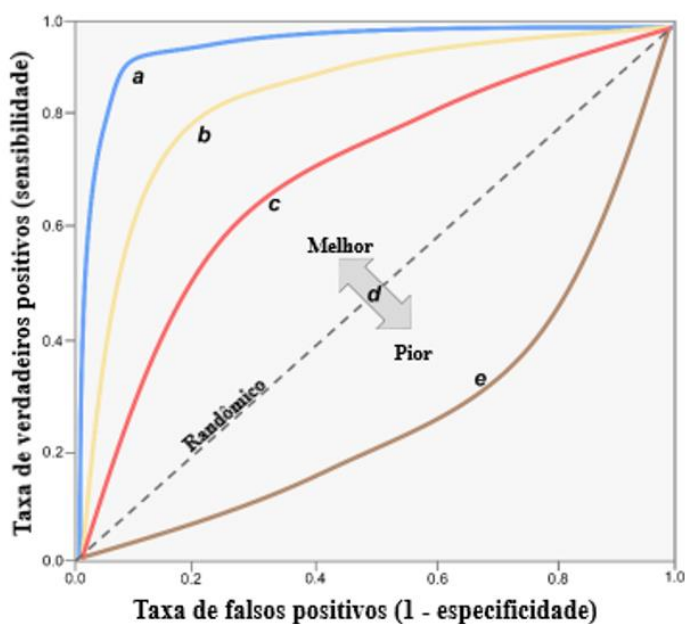


Figura 15 - Diagrama esquemático relacionando o desempenho do acoplamento molecular à curva ROC. A classificação a taxa na curva *a* é melhor do que *b*, *c* e *e*, respectivamente. A curva ROC para um encaixe aleatório a classificação é representada como uma linha pontilhada em *d* (Adaptado de NEVES et al., 2021).

AUC é a probabilidade de um modelo classificar aleatoriamente uma molécula ativa primeiramente em relação a uma molécula inativa (NEVES et al., 2021). A AUC é classificada de acordo com seus resultados, sendo: $AUC < 0.5$ é uma falha; $0.5 \leq AUC < 0.7$ é uma AUC ruim; $0.7 \leq AUC < 0.8$ é aceitável; $0.8 \leq AUC < 0.9$ é bom; e

$0.9 \leq AUC \leq 1$ é excelente. A AUC é calculada conforme a equação 1, representada a seguir:

$$AUC = \frac{1}{n} \sum_{i=1}^n (1 - f_i) \quad (\text{Eq. 1})$$

em que f_i é a fração de *decoys* com classificação superior aos i th ativos.

O fator de enriquecimento (EF) indica o quão melhor é o modelo em ordenar os compostos (ativos e *decoys*) quando comparado a uma lista ordenada aleatoriamente. O cálculo é feito a partir da fração de compostos ativos encontrados no topo da lista ordenada na seleção de $x\%$ do conjunto de dados, a qual é comparada com a fração de compostos ativos encontrados na lista ordenada contendo todos os compostos (ativos e inativos) (HECKER et al., 2002; JACOBSSON et al., 2003) O cálculo do EF pode ser definido pela equação 2:

$$EF_{x\%} = \frac{n_{x\%}/N_{x\%}}{n/N} \quad (\text{Eq. 2})$$

Em que o n representa o número total de ativos, N representa o número total de compostos (ativos e *decoys*) no banco de dados, $n_{x\%}$ representa o número de ativos, e $N_{x\%}$ representa o número total de compostos na lista ordenada de $x\%$, respectivamente (NEVES et al., 2021). No top 1% da lista, o valor máximo de EF é 100, no top 5% o máximo é 20, e no top 10% o valor máximo é 10.

A discriminação aprimorada de Boltzmann da curva ROC (BEDROC, do inglês, *Boltzmann-Enhanced Discrimination of ROC*) (Eq. 4) é uma forma generalizada da AUC com decréscimo exponencial de uma função ponderada que foca em como os compostos ativos são ordenados no início da lista. Em outras palavras, a BEDROC pode ser interpretada como a probabilidade de selecionar um composto ativo antes daqueles provenientes da seleção aleatória de compostos exponencialmente distribuídos para valores de α (TRUCHON; BAYLY, 2007).

$$RIE = \frac{\frac{1}{n} \sum_{i=1}^n e^{-\alpha x_i}}{\frac{1}{n} \left(\frac{1 - e^{-\alpha}}{e^{\frac{\alpha}{n}} - 1} \right)} \quad (\text{Eq. 3})$$

$$BEDROC = RIE \times \frac{R_a \sinh\left(\frac{\alpha}{2}\right)}{\cosh\left(\frac{\alpha}{2}\right) - \cosh\left(\frac{\alpha}{2} - \alpha R_a\right)} + \frac{1}{1 - \exp(\alpha(1 - R_a))} \quad (\text{Eq. 4})$$

A BEDROC é calculada a partir do RIE (Eq. 3), em que RIE corresponde ao enriquecimento inicial robusto, R_a à fração de ativos e α ao parâmetro que controla a distribuição exponencial (BRAGA; ANDRADE, 2013).

4.1.12. Triagem Virtual

Os melhores protocolos e modelos, constatados após avaliação das taxas de enriquecimento, foram incorporados como filtros moleculares na VS de ~1,3 milhões de compostos disponíveis no banco de dados ChemBridge (<https://www.chembridge.com/>).

Com o objetivo de priorizar compostos com maior potencial de inibição contra a POPTc80, inicialmente, todos os compostos (1.360.966) disponibilizados pela ChemBridge foram preparados da mesma forma que os compostos do conjunto de validação (ativos e *decoys*). Em seguida, os compostos foram submetidos sequencialmente a filtragem utilizando o melhor modelo *shape-based* e melhor protocolo de ancoramento molecular (módulos HTVS, SP e XP), respectivamente, desenvolvidos neste trabalho. Ao final desta etapa, os melhores *hits* virtuais foram ainda submetidos a análise e remoção de agregadores (IRWIN et al., 2015) e PAINS (do inglês, *Pan-Assay Interference Compounds*) (BAELL; HOLLOWAY, 2010), utilizando a ferramenta SwissADME (DAINA; MICHIELIN; ZOETE, 2017). Como etapa final de filtragem, uma inspeção visual foi realizada considerando a sobreposição dos *hits* virtuais ao modelo *shape-based*, poses de ancoramento molecular, assim como propriedades físico-químicas (LogP <5).

4.2. Métodos experimentais

4.2.1. Ensaio enzimáticos

Os *hits* virtuais mais promissores identificados a partir da VS serão adquiridos e submetidos a validação experimental *in vitro* para determinação do potencial de inibição de POPTc80. Os ensaios experimentais serão conduzidos em colaboração com a Dra. Izabela Marques Dourado Bastos Charneau Departamento de Biologia Celular da Universidade de Brasília. O ensaio será realizado utilizando o substrato fluorescente N-Suc-Gly-Pro-Leu-Gly-Pro-AMC. A POPTc80 será incubada por 15 min com concentrações seriadas do composto e em seguida a atividade remanescente será obtida pela hidrólise do substrato em placa de 96 poços. A liberação de AMC do substrato será detectada sob 420nmex/460nmem na função cinética (GRELLIER et al., 2001) em espectrofluorímetro (SpectraMax m5, Molecular Devices Inc.). Os resultados serão representados por valores de IC₅₀ e/ou Ki.

5. RESULTADOS

5.1. Predição da estrutura 3D da POPTc80

Face a ausência de estruturas 3D resolvidas experimentalmente e depositadas na base de dados PDB, a estrutura POPTc80 (TcCLB.506247.230) foi predita utilizando três métodos diferentes: modelagem por homologia (SWISS-MODEL), *threading* (I-TASSER) e modelagem *ab initio* baseada em inteligência artificial (AlphaFold). A estrutura predita utilizando o servidor SWISS-MODEL foi predita utilizando como molde a estrutura Prolil oligopeptidase do organismo *Sus scrofa* (código PDB: 4AMY, resolução 2,0 Å, identidade sequencial 42,3%) (KASZUBA et al., 2012). As estruturas 3D geradas pelos servidores I-TASSER e AlphaFold foram utilizadas utilizando abordagens de *threading* e modelagem *ab initio*, respectivamente, as quais são independentes de moldes. As estruturas 3D dos modelos iniciais gerados estão representadas na Figura 15.

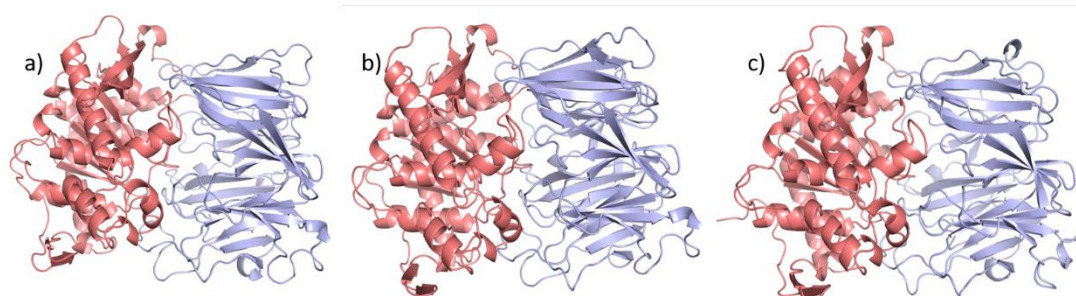


Figura 16 – Estrutura 3D da POPTc80 gerada pelos servidores SWISS-MODEL (a), I-TASSER (b) e AlphaFold (c). Domínio α/β hidrolase e domínio β -propeller representado em salmão e azul claro, respectivamente.

Em seguida, um conjunto de parâmetros geométricos (qualidade dos ângulos, rotâmeros e ligações, regiões Ramachandran favoráveis e desfavoráveis) e escores (*Clashscore* e *MolProbity score*) foram calculados no servidor MolProbity (WILLIAMS et al., 2018) com intuito de estimar a qualidade estrutural dos modelos de POPTc80 (Tabela 4). Todas as estruturas apresentaram *Clashscore* entre 5,52 – 0,92, *MolProbity score* entre 2,81 – 0,89, além de 81,58% a 97,4% dos ângulos de torção ϕ (phi) e ψ (psi) dentro de regiões favoráveis de Ramachandran.

De acordo com a Tabela 4, a estrutura 3D de POPTc80 predita pelo AlphaFold apresentou os melhores resultados (*MolProbity score* 0,89; *Clashscore* 0,92; Regiões Ramachandran favoráveis 97,4%; Regiões Ramachandran desfavoráveis 0,43%). O

gráfico de Ramachandran é um gráfico que expressa todas as combinações de ângulos diedros de aminoácidos de uma macromolécula, permitindo determinar a provável estrutura secundária da biomolécula e a conformação da proteína. O *Clashscore* é o número de choques a cada 1000 átomos. Já o MolProbity score fornece um número que representa estatisticamente a qualidade da proteína, combinando o *Clashscore*, porcentagem de Ramachandran favorecido e não favorecido, e porcentagem de rotâmeros laterais ruins, fornecendo um número a resolução cristalográfica na qual esses valores seriam esperados (WILLIAMS et al., 2018).

Tabela 4 – Validação dos modelos 3D gerados pelo SWISS-MODEL, I-TASSER e AlphaFold, respectivamente.

Parâmetros de validação	SWISS-MODEL	I-TASSER	AlphaFold
<i>Clashscore</i>	2,04	5,52	0,92
<i>MolProbity score</i>	1,44	2,81	0,89
Rotâmeros ruins	7 (1,18%)	67 (11,15%)	4 (0,67%)
Regiões desfavoráveis (Ramachandram)	12 (1,74%)	39 (5,61%)	3 (0,43%)
Regiões favoráveis (Ramachandram)	646 (93,12%)	567 (81,58%)	677 (97,4%)
Ligações ruins	3 (0,05%)	3 (0,05%)	0 (0,00%)
Ângulos ruins	49 (0,64%)	100 (1,3%)	14 (0,18%)

Especificações para a obtenção de estruturas de qualidade: Rotâmeros ruins (<1%), Regiões desfavoráveis (<0,05%), Regiões favoráveis (>98%), Ligações ruins (0%) e Ângulos ruins (<0,1%).

Com o intuito de promover a otimização estrutural dos modelos iniciais gerados, as estruturas 3D de POPTc80 foram refinadas no servidor GalaxyWEB (KO et al., 2012). Os parâmetros geométricos dos modelos refinados, calculados no servidor MolProbity, estão disponíveis na Tabela 4 e 5.

Tabela 5 - Validação dos modelos 3D gerados pelo SWISS-MODEL, I-TASSER e AlphaFold após o refinamento no servidor GalaxyWEB.

Parâmetros de validação	SWISS-MODEL	I-TASSER	AlphaFold
<i>Clashscore</i>	8,63	9,74	5,74
<i>MolProbity score</i>	1,61	1,97	1,32
Rotâmeros ruins	2 (0,34%)	6 (1,00%)	5 (0,83%)

Regiões desfavoráveis (Ramachandram)	1 (0,15%)	10 (1,44%)	0 (0%)
Regiões favoráveis (Ramachandram)	670 (97,2%)	645 (92,8%)	683 (98,27%)
Ligações ruins	14 (0,25%)	21 (0,37%)	18 (0,32%)
Ângulos ruins	87 (1,14%)	97 (1,26%)	68 (0,88%)

Especificações para a obtenção de estruturas de qualidade: Rotâmeros ruins (<1%), Regiões desfavoráveis (<0,05%), Regiões favoráveis (>98%), Ligações ruins (0%) e Ângulos ruins (<0,1%).

De acordo com a Tabela 5, o refinamento estrutural resultou em melhoras significativas na qualidade geométrica das estruturas obtidas nos servidores SWISS-MODEL (regiões Ramachandran favoráveis = ↑4,08 %; regiões Ramachandran desfavoráveis = ↑1,59%; *MolProbity score* = 1,61; *Clashscore* = 8,63) e I-TASSER (regiões Ramachandran favoráveis = ↑11,22%; regiões Ramachandran desfavoráveis = ↑4,17%; *MolProbity score* = 1,97; *Clashscore* = 9,74). Por outro lado, o refinamento estrutural não resultou em melhorias significativas na qualidade estrutural do modelo obtido no AlphaFold. Diante destes resultados, a estrutura 3D gerada pelo AlphaFold, sem refinamento estrutural, foi considerada como melhor modelo e, portanto, foi utilizada nas análises descritas nas seções abaixo.

5.2. Análise estrutural da POPTc80

O melhor modelo 3D de POPTc80 foi submetido ao processo de geração de conformações no servidor *DynOmics* utilizando a técnica de *Elastic Network Models* (ENM). Através desta abordagem, foram geradas 29 conformações POPTc80. Em seguida, três conformações de POPTc80 representativas (Conformação 09,15 e 20) foram selecionadas através de uma PCA (Figura 16 a - c) e predição de bolsos de interação no servidor FTSite (Figuras 17 - 19).

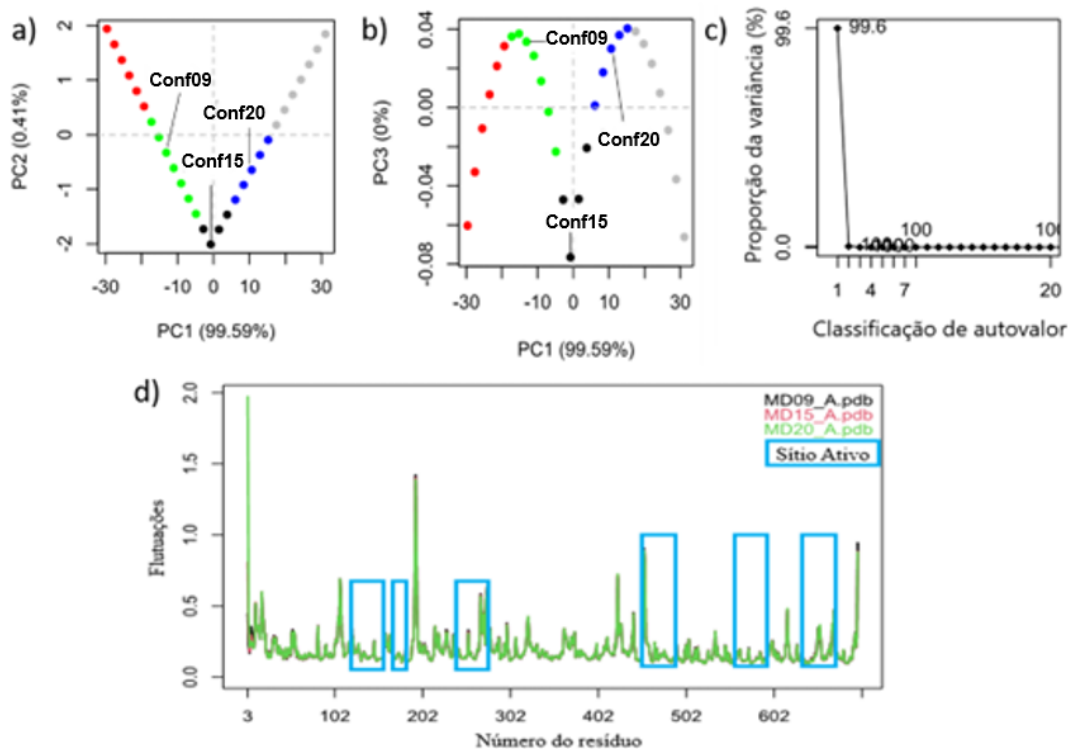


Figura 17 – Análise de PCA da POPTc80 e suas flutuações. Análise de PCA do modelo gerado pelo AlphaFold. Comparação entre PC1 e PC2 (a); PC1 e PC3 (b) e a proporção da variância relacionada à classificação de autovalor (c) e gráfico de flutuações (d). Cada ponto representa uma estrutura e a cor do ponto indica o ID de cluster de um cluster conformacional.

De acordo com a Figura 16d, as conformações 1,2 e 3 de POPTc80 apresentam baixa flutuação, com valores de RMSD $<1,0$ Å nas regiões correspondentes ao bolso de ligação de substratos peptídicos (Resíduos: 115-160, 170-190, 240-290, 460-490, 540-595, 630-685, aproximadamente). Por outro lado, a análise de *hotspots* para predição dos bolsos de interação (Figura 18) demonstrou que estas flutuações conformacionais podem afetar significativamente a ligação de inibidores ao sítio ativo de POPTc80.

Conforme descrito na Figura 18, dois *hotspots* de ligação foram preditos em regiões adjacentes ao domínio α/β hidrolase da conformação 9. Por outro lado, a conformação Conformação 15 apresenta um *hotspot* localizado em região isolada do domínio β -propeller. A conformação 20 por sua vez apresenta *hotspots* 1 e 2 em regiões adjacentes ao domínio α/β hidrolase, enquanto que um terceiro *hotspot* foi identificado próximo ao *hotspot* 2.

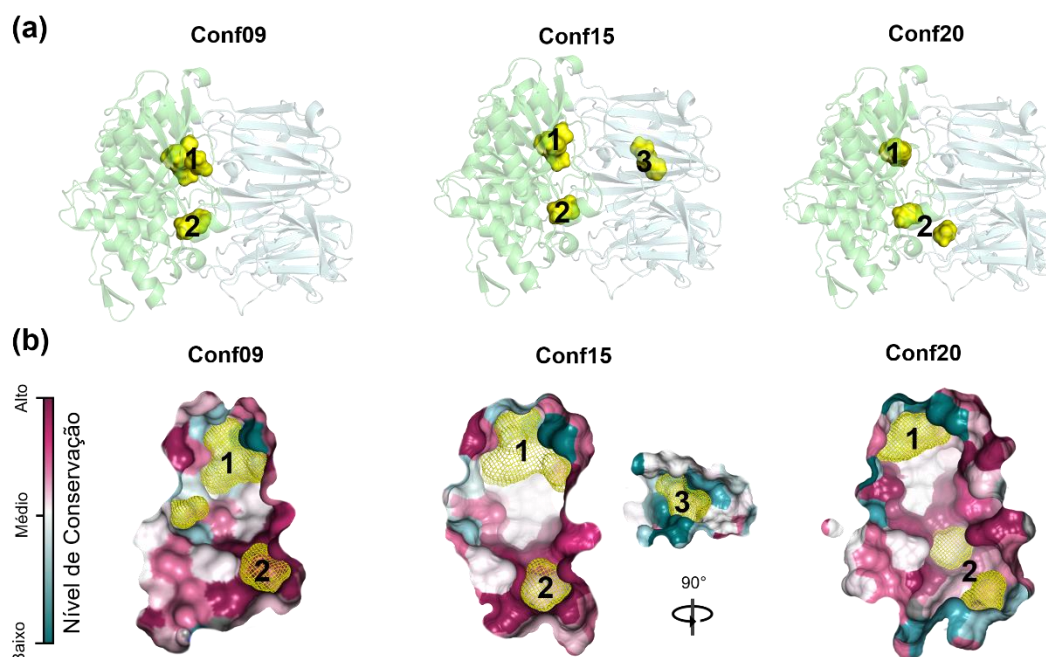


Figura 18 – Predição de sítios de ligação nas conformações Conf09, Conf15 e Conf20 de POPTc80 utilizando os servidores FTSite (a) e análise de conservação dos resíduos de aminoácido baseada no servidor ConSurf (b).

Após predição dos sítios de ligação, a estrutura 3D de POPTc80 foi submetida ao servidor ConSurf (ASHKENAZY et al., 2016) com vistas a avaliar o grau de importância estrutural dos aminoácidos próximos aos *hotspots*. O servidor ConSurf realiza uma análise que permite revelar regiões funcionais em macromoléculas por meio da análise da dinâmica evolutiva das substituições de aminoácidos entre sequências homólogas (ASHKENAZY et al., 2016).

A análise ConSurf demonstrou que os resíduos Ile66, Gln70, Tyr465, Ser479, Ser480, Ser481, Arg482, Met545, Gln571, Gly668, Lys671, Lys675, Ile676, Leu677, Glu679, Ala680, Arg682, Ile683, Tyr684, próximos ao *hotspot* 1 das conformações 09, 15 e 20 apresentam baixo grau de conservação estrutural. Por outro lado, a região correspondente ao *hotspot* 2 possui resíduos de aminoácidos (Tyr467, Phe470, Ser548, Asn549, Val574, Trp589, Tyr593, Arg633, Val634, His667) altamente conservados. Notoriamente, os resíduos Ser548, Asp631 e His667 conservados, próximos ao *hotspot* 2, correspondem aos aminoácidos da tríade catalítica (BASTOS et al., 2005). Nesta mesma região, também foi identificado que os resíduos conservados Trp589 e Arg633 correspondem aos resíduos responsáveis pela interação com inibidores competitivos (co-cristalizados)

(6S)-1-chloro-3-[(4-fluorobenzyl)oxy]-6-(pyrrolidin-1-ylcarbonyl)pyrrolo[1,2-a]pyrazin-4(6H)-one (2); 1-[(2R,5S)-2-terc-butil-5-((2S)-2-

[(1R)-1,2-dihidroxietil]pirrolidin-1-il]carbonil]pirrolidin-1-il]-4-fenilbutan-1-ona (3); 1-[4-oxidanil-2-(1,3-tiazolidin-3-ilcarbonil)pirrolidin-1-il]-4-[2-[(~{E})]-2-feniletetil]fenoxi]butan-1-ona (4); disponíveis nas estruturas de POP de *Homo sapiens*, *Sus scrofa* e *Haliotis discus hannai* (com os códigos PDBs: 3DDU, 4AMY e 6JCI, respectivamente) (Figura 19). Estes resultados indicam que os resíduos conservados Trp589 e Arg633, próximos ao *hotspot* 2, podem desempenhar um papel funcional evolutivo essencial para modulação da atividade catalítica ou promoção de inibição da POPTc80.

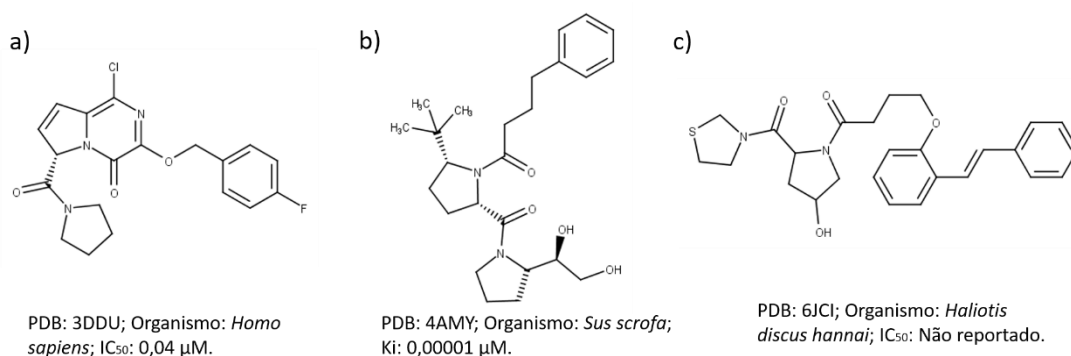


Figura 19 – Estrutura 2D dos inibidores co-cristalizados de prolil oligopeptidases e seus valores de inibição biológico.

Diante destes resultados, foi proposto um sítio de ligação preferencial considerando a análise de *hotspots* e grau de conservação dos resíduos de aminoácido na conformação 20. Um modelo hipotético do modo de ligação de inibidores de POPTc80 está representado na Figura 20, onde as interações com os resíduos Arg633 e Trp589 são essenciais para que ocorra um alto grau de interação entre o ligante e a proteína. Além disso, as interações hidrofóbicas nos bolsos S1', S1, S2 e S3 também são importantes (KUMAR et al., 2017).

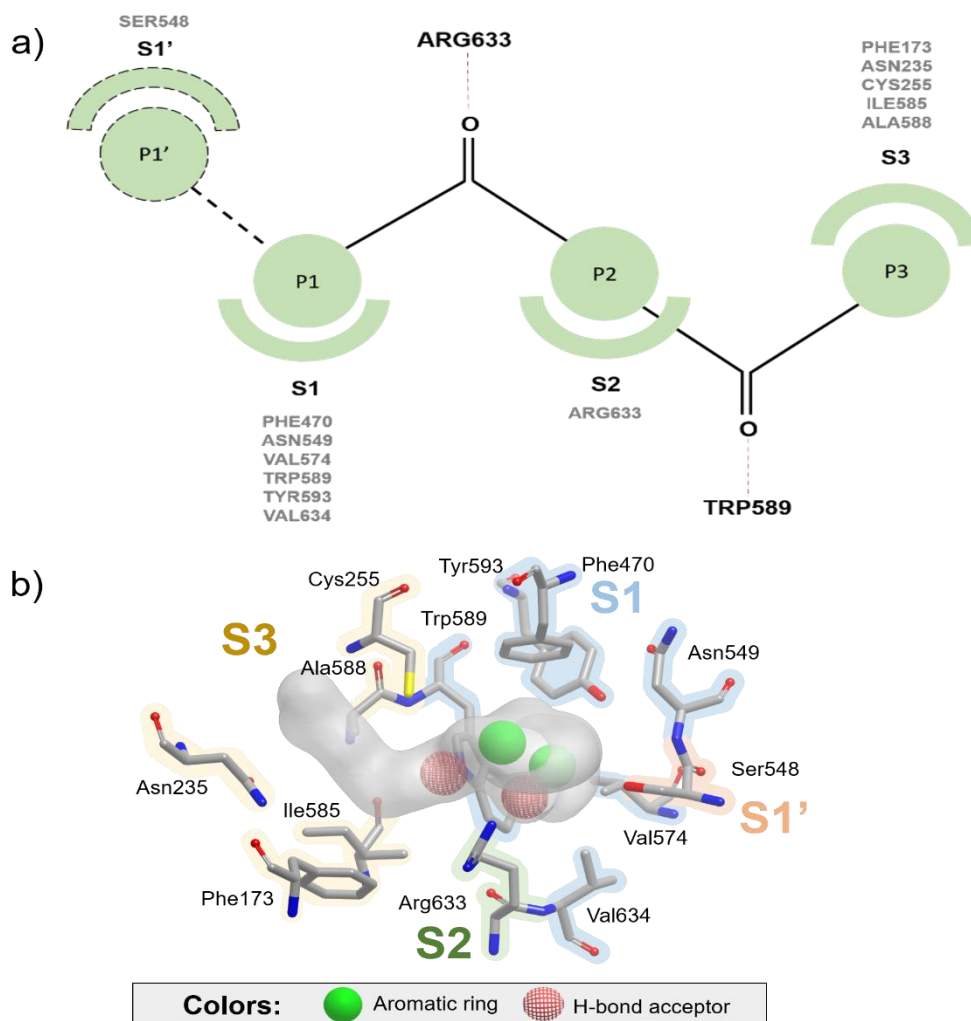


Figura 20 – Modelo hipotético do modo de ligação de inibidores de POPTc80 (a) e o posicionamento da *Query 4* em relação ao sítio da POPTc80 (b).

5.3. Validação do ancoramento molecular

Estudos de ancoramento molecular foram realizados para as conformações 09, 15 e 20 de POPTc80 utilizando os módulos HTVS, SP e XP do programa Glide e as funções de pontuação *Docking Score*, *Glide Emodel* e o *Glide Energy*. As análises de ancoramento molecular foram conduzidas em dois sítios de ligação independentes (sítios 1 e 2) identificados na interface dos domínios α/β hidrolase e β -propeller. Os resultados estatísticos de validação dos protocolos de ancoramento molecular estão disponíveis nas Tabela 6-11.

Conforme apontado nas Tabelas 6-7, o ranqueamento do conjunto de validação (ativos e *decoys*) utilizando a função de pontuação empírica *Docking Score* não resultou em taxas de enriquecimento satisfatórias (EF $\sim 0,0$) e BEDROCs $\sim 0,0$). Diante

20 0,81 0,00 0,00 0,00 0,00 0,00 0,01

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. HTVS: modo de triagem virtual de alto rendimento; SP: modo de precisão padrão; XP: modo de precisão extra.

Considerando a classificação aleatória de ativos e *decoys* utilizando a função de pontuação *Docking Score*, outras duas funções de pontuação (*Glide Emodel* e *Glide Energy*) baseadas nos termos de energia do escore do *grid*, afinidade de ligação predita pelo *GlideScore* e energia de interação de Coulomb-van der Waals modificada foram exploradas nos estudos de ancoramento molecular. De acordo com a Tabela 8, os protocolos de ancoramento molecular no sítio 1 utilizando a função de pontuação *Glide Emodel* levaram a taxas de enriquecimento satisfatórias nos módulos SP (AUC = 0,71 – 0,73, EF_{1%} = 18,5 – 26,9 e BEDROC_{1%} = 0,29 – 0,43) e XP (AUC = 0,74-0,77, EF_{1%} = 19,2-25,9 e BEDROC_{1%} = 0,34-0,39). Apesar disso, não foi possível identificar a partir desta análise qual seria a conformação preferencial para a ligação de compostos ativos.

Tabela 8 - Resultado da validação dos protocolos de ancoramento molecular no sítio 1 da enzima POPTc80 utilizando a função de pontuação *Glide Emodel*.

Conformação	AUC	Top 1%		Top 5%		Top 10%	
		EF	BEDROC	EF	BEDROC	EF	BEDROC
Módulo HTVS							
09	0,63	14,9	0,33	5,93	0,29	4,44	0,33
15	0,66	22,2	0,42	5,93	0,33	3,70	0,34
20	0,63	23,1	0,39	6,92	0,33	3,85	0,35
Módulo SP							
09	0,72	25,9	0,40	8,15	0,39	4,56	0,42
15	0,73	18,5	0,29	8,15	0,36	4,81	0,41
20	0,71	26,9	0,43	9,23	0,39	4,62	0,44
Módulo XP							
09	0,77	22,2	0,34	8,89	0,39	5,56	0,45
15	0,74	25,9	0,39	8,15	0,38	4,81	0,43
20	0,74	19,2	0,37	7,69	0,38	5,00	0,43

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. HTVS: modo de triagem virtual de alto rendimento; SP: modo de precisão padrão; XP: modo de precisão extra. O melhor modelo está destacado em negrito.

Por outro lado, o protocolo de ancoramento molecular no sítio 2 utilizando a função de pontuação *Glide Emodel* (Tabela 9) apresentou taxas de enriquecimento superiores aquelas demonstradas anteriormente para o sítio 1. De acordo com a Tabela 7 os protocolos de ancoramento na conformação 20 resultaram nas melhores taxas de enriquecimento nos módulos HTVS (AUC = 0,82, EF_{1%} = 26,9 e BEDROC_{1%} = 0,53), SP (AUC = 0,84, EF_{1%} = 30,8 e BEDROC_{1%} = 0,51) e XP (AUC = 0,84, EF_{1%} = 26,9 e BEDROC_{1%} = 0,52). Apesar disso, não foi possível identificar a partir desta análise qual seria a conformação preferencial para a ligação de compostos ativos. Estes resultados corroboram com as análises estruturais exploradas nas seções anteriores, demonstrando que o sítio 2 é potencialmente o bolso de interação preferencial para inibidores de POPTc80.

Tabela 9 – Resultado da validação dos protocolos de ancoramento molecular no sítio 2 da enzima POPTc80 utilizando a função de pontuação *Glide Emodel*.

Conformação	AUC	Top 1%		Top 5%		Top 10%	
		EF	BEDROC	EF	BEDROC	EF	BEDROC
Módulo HTVS							
09	0,68	22,2	0,45	5,19	0,32	3,70	0,34
15	0,76	22,2	0,45	5,93	0,31	4,44	0,36
20	0,82	26,9	0,53	10,8	0,48	5,77	0,52
Módulo SP							
09	0,89	25,9	0,51	10,4	0,47	5,56	0,52
15	0,92	29,6	0,55	11,1	0,49	6,67	0,56
20	0,91	30,8	0,51	11,5	0,50	6,54	0,57
Módulo XP							
09	0,82	25,9	0,46	8,15	0,41	5,19	0,47
15	0,83	29,6	0,47	7,41	0,41	5,41	0,47
20	0,84	26,9	0,52	9,23	0,44	5,38	0,49

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. HTVS: modo de triagem virtual de alto rendimento; SP: modo de precisão padrão; XP: modo de precisão extra. Os melhores protocolos gerados estão destacados em negrito.

Parâmetros de validação estatística semelhantes aqueles obtidos para a função de pontuação *Glide Emodel* (Tabelas 8 e 9) também foram obtidos para a função de pontuação *Glide Energy*. De acordo com a Tabela 11, o ancoramento molecular na conformação 20 resultou nas melhores taxas de enriquecimento para os módulos

HTVS (AUC = 0,88, EF_{1%} = 23,1 e BEDROC_{1%} = 0,48) e XP (AUC = 0,86, EF_{1%} = 26,9 e BEDROC_{1%} = 0,56), enquanto que o ancoramento na conformação 15 resultou na melhor taxa de enriquecimento para o módulo SP (AUC = 0,87, EF_{1%} = 27,2 e BEDROC_{1%} = 0,56).

Tabela 10 – Resultado da validação dos protocolos de ancoramento molecular no sítio 1 da enzima POPTc80 utilizando a função de pontuação *Glide Energy*.

Conformação	AUC	Top 1%		Top 5%		Top 10%	
		EF	BEDROC	EF	BEDROC	EF	BEDROC
Módulo HTVS							
09	0,65	18,5	0,36	7,41	0,34	4,44	0,39
15	0,61	22,2	0,42	7,41	0,35	4,07	0,37
20	0,69	23,1	0,38	6,15	0,33	4,23	0,37
Módulo SP							
09	0,73	29,6	0,45	8,89	0,42	4,81	0,45
15	0,73	25,9	0,36	8,15	0,40	5,19	0,45
20	0,73	30,8	0,45	9,23	0,44	5,38	0,48
Módulo XP							
09	0,78	18,5	0,35	8,89	0,39	5,56	0,46
15	0,77	22,2	0,40	10,4	0,43	5,93	0,49
20	0,75	28,2	0,45	10,0	0,47	5,77	0,51

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. HTVS: modo de triagem virtual de alto rendimento; SP: modo de precisão padrão; XP: modo de precisão extra.

Tabela 11 – Resultado da validação dos protocolos de ancoramento molecular no sítio 2 da enzima POPTc80 utilizando a função de pontuação *Glide Energy*.

Conformação	AUC	Top 1%		Top 5%		Top 10%	
		EF	BEDROC	EF	BEDROC	EF	BEDROC
Módulo HTVS							
09	0,69	22,2	0,41	7,41	0,32	4,44	0,36
15	0,77	18,5	0,42	8,89	0,36	5,19	0,42
20	0,81	23,1	0,53	10,0	0,47	5,77	0,52
Módulo SP							
09	0,87	27,2	0,53	9,63	0,47	5,56	0,51
15	0,90	25,9	0,50	8,15	0,42	6,15	0,49

20	0,88	23,1	0,48	10,0	0,45	5,38	0,50
Módulo XP							
09	0,85	25,9	0,49	8,89	0,44	5,56	0,49
15	0,88	18,5	0,39	7,41	0,35	5,19	0,43
20	0,86	26,9	0,56	8,46	0,46	5,00	0,49

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. HTVS: modo de triagem virtual de alto rendimento; SP: modo de precisão padrão; XP: modo de precisão extra. Os melhores protocolos gerados estão destacados em negrito.

5.4. Modelos *Shape-Based*

Modelos de forma e molecular foram construídos para distinguir compostos potencialmente ativos para a POPTc80 utilizando como arquivos de entrada a poses de ancoramento de 6 inibidores conhecidos (*queries*, conferir Tabela 6) no sítio 1 da

conformação 20 de POPTc80. Os diagramas de interação 2D e modelos de forma e volume molecular gerados estão destacados na Figuras 21 e 22.

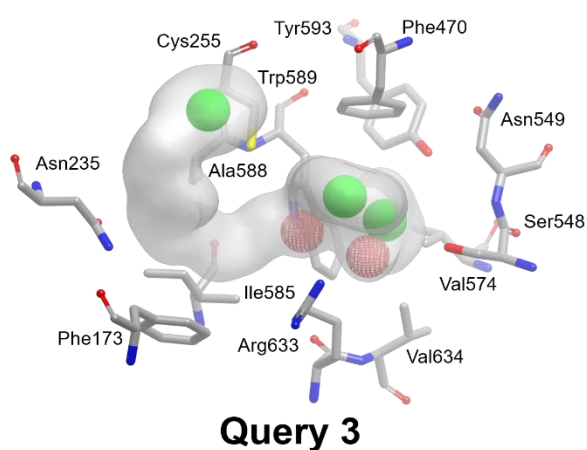
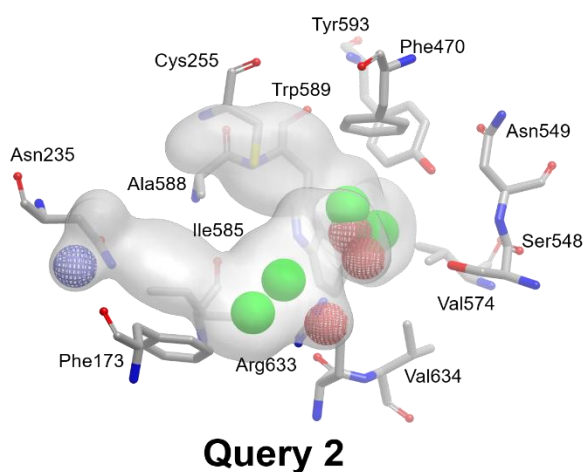
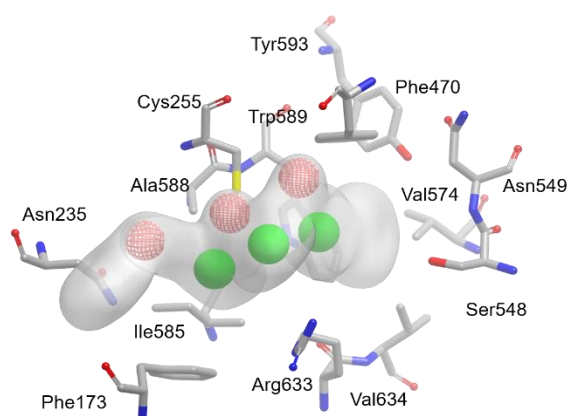


Figura 21 – Diagrama 3D das *queries* 1, 2 e 3 dos modelos *Shape-based*, sendo as cores: doador (azul tracejado), aceitador (vermelho tracejado) e anel (verde).

A maior parte das *queries* faz ligação de hidrogênio com os aminoácidos Arg633 e Trp569, assim como interações de van der Waals com os resíduos de aminoácidos: Ser548 (bolso S1'); Phe70, Asn549, Val574, Trp589, Tyr593, Val634 (bolso S1); Arg633 (bolso S2); e Phe173, Asn235, Cys255, Ile585, Ala588 (bolso S3). Este padrão de interação intermolecular foi considerado como base estrutural para definição das cores (características farmacofóricas) utilizadas nos modelos de forma e volume molecular.

Conforme observado na Tabela 12, todos os modelos de forma e volume molecular desenvolvidos apresentaram valores de AUC entre 0,78 – 0,92. O modelo gerado com a *query* 4 apresentou o melhor desempenho estatístico superior, com valores de EF de 48,1, 13,3, e 7,04; e valores de BEDROC de 0,80, 0,71, e 0,72 nos top 1%, 5% e 10% da lista de classificação, respectivamente. Este modelo foi gerado com as seguintes cores: doador (azul tracejado), aceitador (vermelho tracejado) e anel (verde). A cor “doador” representa regiões onde há a presença de receptores de hidrogênio, como grupamentos amina. A cor “aceitador” representa regiões onde há a presença de aceptores de hidrogênio, como átomos de oxigênio. A cor “anel” representa a presença de anéis de 5 ou 6 membros.

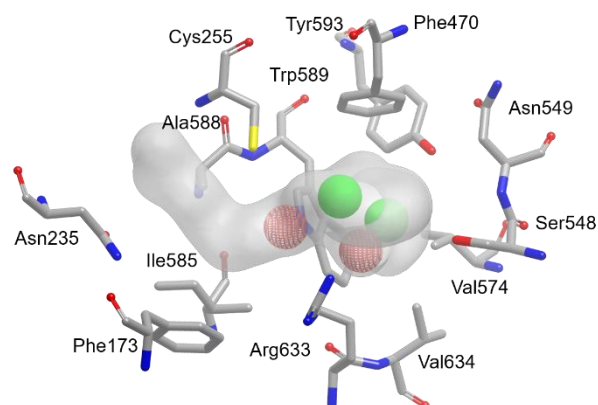
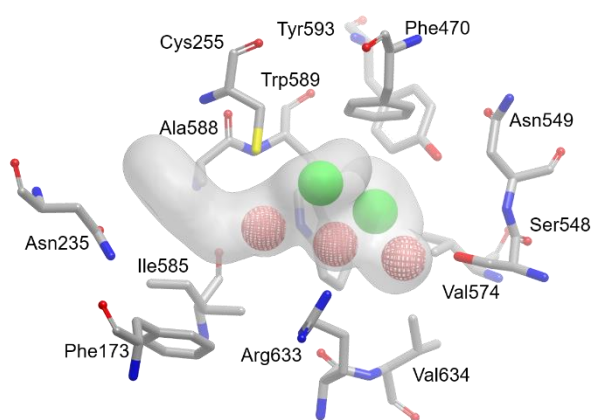
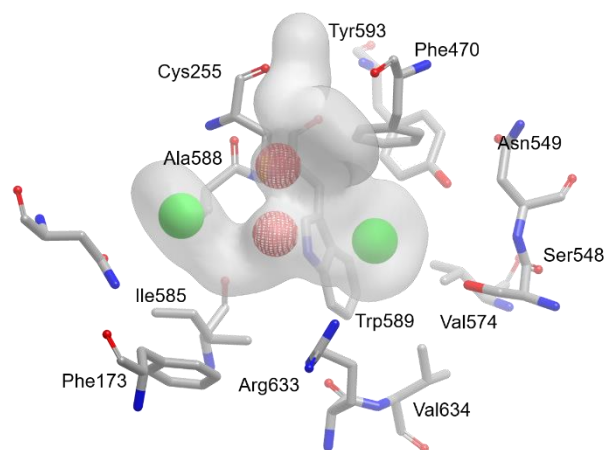
**Query 4****Query 5****Query 6**

Figura 22 – Diagrama 3D das *queries* 4, 5 e 6 dos modelos *Shape-based*, sendo as cores: aceitador (vermelho tracejado) e anel (verde).

Tabela 12 - Desempenho estatístico dos modelos de *Shape-Based* desenvolvidos neste trabalho.

Modelos	AUC	Top 1%		Top 5%		Top 10%	
		EF	BEDROC	EF	BEDROC	EF	BEDROC
<i>Query 1</i>	0,91	25,0	0,44	12,6	0,51	8,15	0,61
<i>Query 2</i>	0,78	22,2	0,41	6,67	0,31	4,07	0,36
<i>Query 3</i>	0,92	33,3	0,42	11,8	0,53	7,01	0,62
<i>Query 4</i>	0,92	48,1	0,82	13,3	0,71	7,04	0,72
<i>Query 5</i>	0,91	39,8	0,71	13,3	0,68	8,15	0,72
<i>Query 6</i>	0,90	25,0	0,37	11,8	0,51	7,78	0,61

AUC: área sob a curva ROC; EF: fator de enriquecimento; BEDROC: discriminação aprimorada de Boltzmann da curva ROC; Top x% *hits*: métricas calculadas para os primeiros x% compostos da lista ordenada pelo score *Reference Tversky Combo*. O melhor modelo *shape-based* está destacado em negrito.

5.5. Triagem Virtual

Primeiramente, todos os compostos (1.360.966) disponibilizados pela ChemBridge foram preparados da mesma forma que os compostos da validação. Os compostos, já preparados, passaram pelo filtro do melhor modelo *Shape-Based* (modelo 4), afim de encontrar novos inibidores da POPTc80 de *T. cruzi*. Em seguida, foi feito um ancoramento molecular com as Top 1% moléculas (13.608) que passaram pelo primeiro filtro. O primeiro ancoramento molecular foi feito no Glide utilizando o modo HTVS, e foi selecionado as melhores 50% das moléculas (6.707) com os melhores resultados do Glide. Logo em seguida, essas moléculas passaram por um novo ancoramento molecular, agora na função SP, e foi retido as melhores 10% das moléculas (672). Depois da função SP, as moléculas passaram pela função XP, que após os cálculos, foi retido também as melhores moléculas no Top 10%, restando 67 moléculas. Por fim, as moléculas passaram por um filtro de agregadores e também um filtro de PAINS, restando 64 moléculas.

Para a seleção destas moléculas, foi feita uma inspeção visual, levando em consideração o seu modo de ligação, *Glide Emodel*, *RefTverskyCombo*, solubilidade e entre outros. Ao final da triagem virtual (Figura 23), os 15 *hits* virtuais foram adquiridos e submetidos a avaliação experimental.

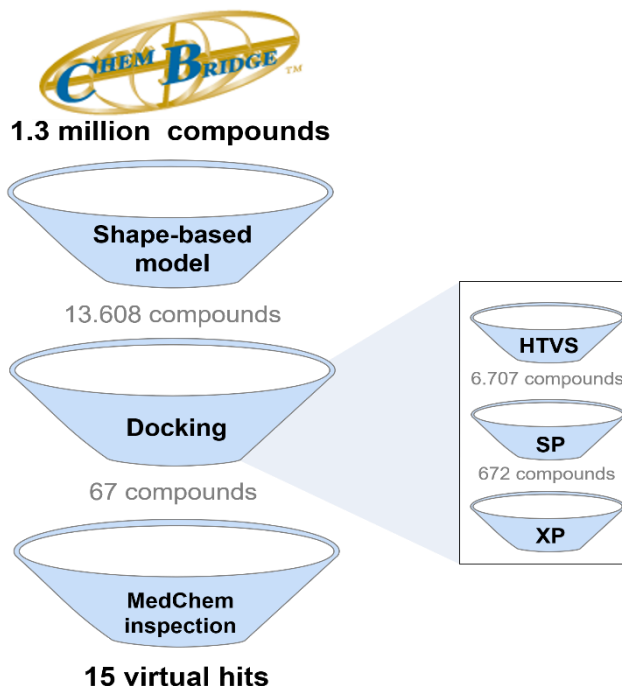


Figura 23 – Etapas de filtragem exploradas na VS de novos inibidores de POPTc80 utilizando a base de dados ChemBridge.

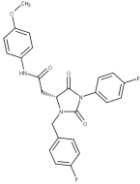
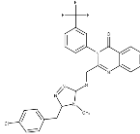
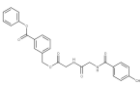
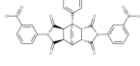
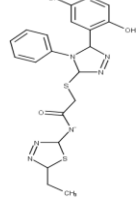
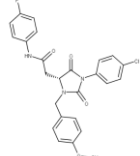
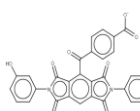
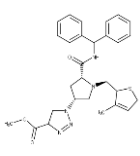
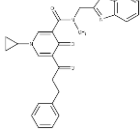
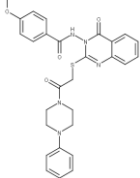
As estruturas químicas dos 15 *hits* virtuais selecionados, assim como escores e propriedades físico-químicas associadas estão representadas na Tabela 13. Esses compostos seguiram para a etapa de validação experimental.

Tabela 13 - *Hits* virtuais selecionados a partir da VS.

ID	Estrutura	RefTversky Combo	Glide Emodel	Peso Molecular	LogP	LogS
7898173		1,514	-96,112	555	3,59	-6,62
5634226		1,407	-88,947	585	3,77	-6,68

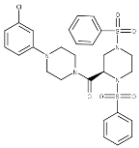
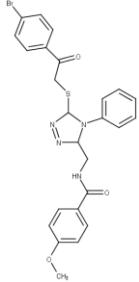
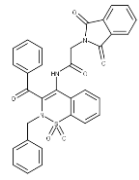
(continua)

Tabela 13 - Hits virtuais selecionados a partir da VS.

ID	Estrutura	RefTversky Combo	Glide Emodel	Peso Molecular	LogP	LogS
7991224		1,282	-88,307	465	4,02	-6,02
7961087		1,299	-87,911	542	4,45	-6,95
7886969		1,264	-86,428	460	3,98	-5,96
5133499		1,411	-86,11	565	2,7	-6,01
7960320		1,305	-84,819	517	3,61	-5,47
7863291		1,282	-83,183	482	4,69	-6,72
5187171		1,306	-82,973	548	2,55	-6,49
25088813		1,299	-81,824	516	4,1	-5
60449174		1,34	-80,133	473	1,47	-6,44
7761043		1,288	-80,01	530	2,04	-5,16

(continua)

Tabela 13 - *Hits* virtuais selecionados a partir da VS.

ID	Estrutura	RefTversky Combo	Glide Emodel	Peso Molecular	LogP	LogS
7620188		1,384	-79,717	589	4,81	-5,76
7784618		1,297	-79,591	537	4,27	-6,76
7691100		1,333	-79,503	578	4,88	-6,69

(conclusão)

Os *hits* virtuais demonstram interações moleculares semelhantes com os inibidores co-cristalizados de outras POPs. O *hit* 7898173 (Figura 24) demonstrou interação com as regiões do sítio (S1, S2 e S3), sendo interação de van der Waals com a Asn235, Phe252, Phe470, Asn471, Ile472 e Val574, interação de empilhamento de anéis com os resíduos Trp589 e Phe470, ligação de hidrogênio com a Arg290, outra ligação de hidrogênio com a Cys255, e ligação iônica com a Arg633.

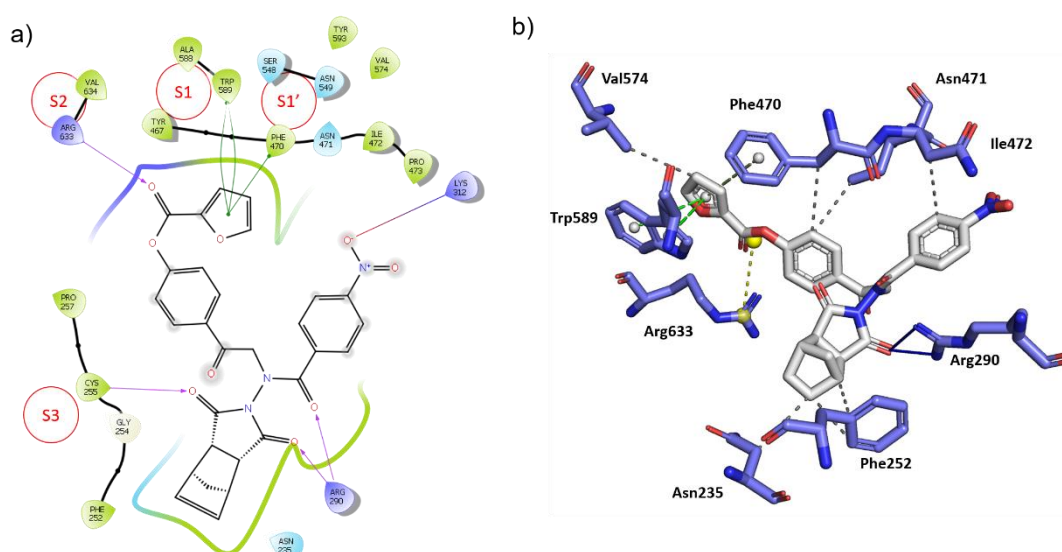


Figura 24 – Interações moleculares do *hit* 7898173. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.

O *hit* 5634226 (Figura 25) mostra interações nas regiões S1, S2 e S3 do sítio, sendo interações eletrostática com os resíduos Phe173, Thr175, Asn235, Phe252, Ile472, Ile585 e Ala588. Interação pi-cátion com a Arg633, empilhamento de anéis com o Phe470 e Phe252, e ligações de hidrogênio com os resíduos Arg290, Trp589 e Asn235.

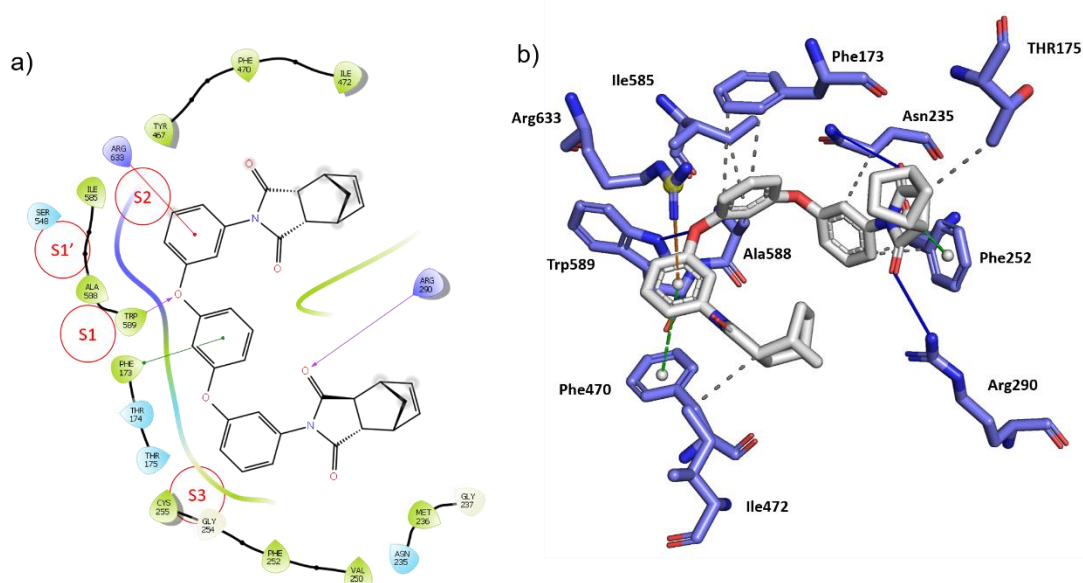


Figura 25 - Interações moleculares do *hit* 5634226. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.

O *hit* 7991224 (Figura 26) também interage com as principais regiões do sítio ativo. Suas interações são interações van der Waals com os resíduos Pro257, Asn471, Ile585, Ala588, Trp589, Arg633 e Val634. Ligações de hidrogênio realizadas pelos resíduos Cys255, Tyr467, Arg633, Trp589 e Asn471. Interações de empilhamento de anéis pelos resíduos Trp589 e Phe470.

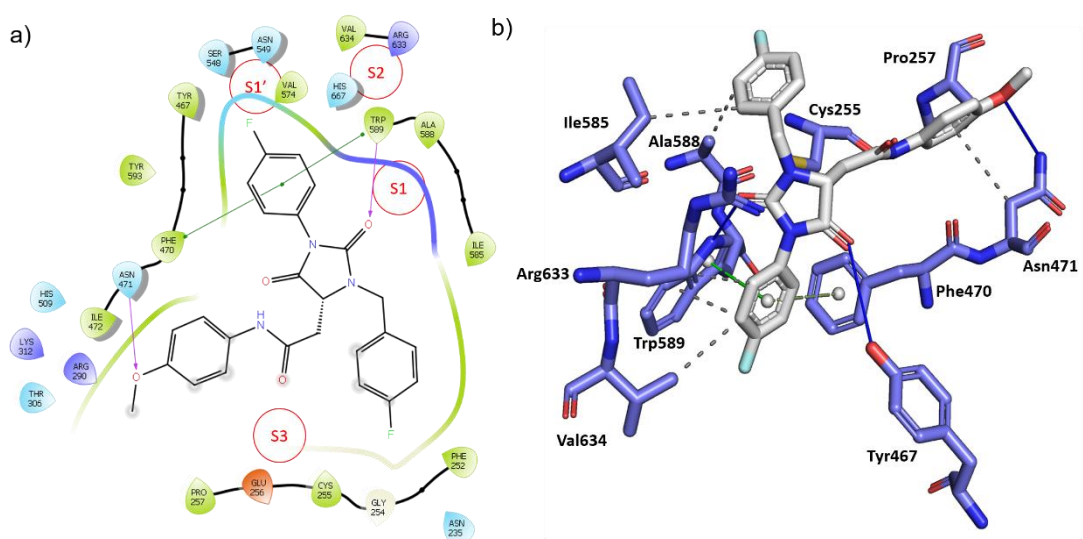


Figura 26 - Interações moleculares do *hit* 7991224. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π -

π stacking (empilhamento de anéis) tracejado em amarelo. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.

Já o *hit* 7961087 (Figura 27) mostrou interações hidrofóbicas com os resíduos Phe252, Trp589, Arg633. As ligações de hidrogênio ocorreram com os resíduos Asn235, Arg290 e Trp589. Interações do tipo empilhamento de anéis foram encontradas com os resíduos Phe252, Phe470 e Trp589. Esse ligante também faz interação do tipo π -cátion com os resíduos Arg290 Arg633.

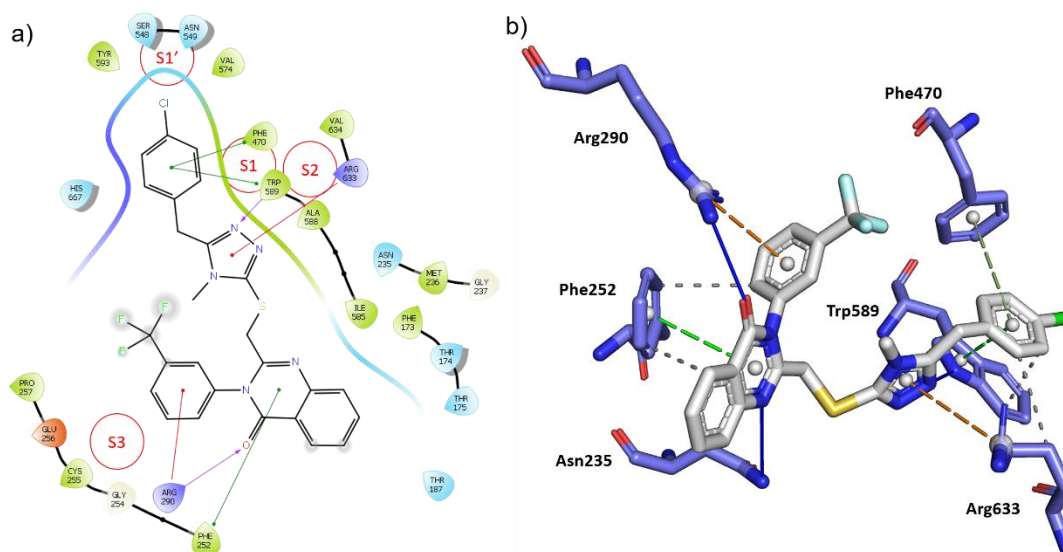


Figura 27 - Interações moleculares do *hit* 7961087. Na esquerda (a), o diagrama de interação mostra as interações intermoleculares. Na direita (b), as interações de hidrogênio estão em azul, ligação iônica em tracejado amarelo e interações do tipo π - π stacking (empilhamento de anéis) em verde tracejado, interações do tipo π -cátion em laranja tracejado. A estrutura 2D teve seu sítio foi mapeado conforme a descrição já mencionada acima.

O *hit* 7886969 (Figura 28) realiza de interação de van der Waals com os resíduos Val334, Leu350, Phe470, Ile472, Val574, Trp589, Arg633 e Val634. O composto também ligações de hidrogênio com os resíduos Phe470 e Asn471. Os empilhamentos de anéis foram apontados nos resíduos Phe470 e Trp589. Além disso, uma interação do tipo π -cátion foi encontrada com o resíduo Lys312. E por fim, ligações iônicas também foram encontradas com os resíduos Arg290 e Arg633.

DISCUSSÃO

Devido ao perfil toxicológico do benznidazol e a sua baixa eficácia no tratamento da doença de Chagas, é necessário a descoberta de novos alvos moleculares e novos fármacos com novos mecanismos de ação. A prolil oligopeptidase 80 é um alvo promissor para a descoberta e desenvolvimento de novos candidatos a fármacos tripanostáticos, por estar envolvida na degradação específica de colágenos humanos purificados de tipo I e IV e a hidrólise nativa de colágeno tipo I, e com a inibição dessa enzima, é possível prejudicar a entrada do parasito nas células hospedeiras (BASTOS et al., 2005; GRELLIER et al., 2001). Esses achados reforçaram a importância da POPTc80 como um alvo bioquímico promissor para o planejamento de novos fármacos tripanostáticos. Sendo assim, os estudos desta dissertação focaram na identificação de inibidores de prolil oligopeptidase 80 de *T. cruzi*, ainda ausentes no mercado.

Nesta dissertação, diversas ferramentas do CADD foram amplamente utilizadas, com o intuito de reduzir tempo e custo no processo de descoberta de novos candidatos a fármacos anti chagásico. A indústria farmacêutica e a academia têm utilizado as abordagens do CADD para reduzir inicialmente o número de compostos em ensaios experimentais (LEELANANDA; LINDERT, 2016).

No estudo dessa dissertação, foi obtido três estruturas 3D da prolil oligopeptidase 80 de *Trypanosoma cruzi*. Essas estruturas foram comparadas e a estrutura do AlphaFold demonstrou resultados mais promissores. O AlphaFold já demonstrou sua utilidade para a comunidade experimental (LEIPART et al., 2022), tanto para substituição molecular quanto para interpretação de microscopia eletrônica criogênica. Além disso, o AlphaFold produz previsões em minutos de unidade de processamento gráfico (GPU) a horas de GPU, dependendo do comprimento da sequência da proteína (JUMPER et al., 2021).

Na análise da drogabilidade dessas estruturas, foi observado que existem dois possíveis sítios de interação de ligantes, que foram chamados de sítio 1 e sítio 2. O sítio 3, encontrado na conformação 15, representa um possível sítio alostérico onde essa região não foi explorada nos estudos de *docking* devido a rigidez do domínio β -propeller na POPTc80. Diante disso, foi realizado um ancoramento molecular em

ambas as regiões com potencial. Nota-se que os resultados do ancoramento molecular no sítio 2 foram mais promissores.

Ao analisar a drogabilidade da POPTc80, tanto o sítio 1 quanto o sítio 2 aparentam ser promissores, e os compostos demonstram uma maior afinidade pelo sítio 1. Entretanto, as informações de homólogos depositados no PDB (*Protein Data Bank*) são mais confiáveis que os resultados de um ancoramento molecular. Diante do exposto, esse estudo foi redirecionado para a busca de novos inibidores desta proteína, através da interação na região do sítio do co-cristal (sítio 2) das estruturas homólogas. Além disso, a análise do servidor *ConSurf* confirma que os aminoácidos da região do ligante co-cristalizado confirmam que esses resíduos se mantêm conservados baseados em estruturas homólogas dessa proteína. A presença de aminoácidos conservados é uma das formas de encontrar resíduos que estejam envolvidos com a interação de ligantes (CAPRA; SINGH, 2007).

De acordo com Katsuno e colaboradores (2015), apoiados nas definições feitas em conjunto por autoridades do GHIT (Global Health Innovative Technology Fund) e do DNDi, o critério para a seleção de *hits* para doença de Chagas se baseia na obtenção de compostos com potência micromolar *in vitro* ($< 10 \mu\text{M}$) contra *T. cruzi* intracelular e seletividade maior que 10 (KATSUNO et al., 2015). Considerando esse critério, nesse trabalho, 24 compostos ativos e 7 inativos foram encontrados na literatura.

Os métodos baseados na forma e volume moleculares não se limitam somente na sobreposição de grupos específicos (doadores e aceptores de ligação de hidrogênio, hidrofóbicos, etc.), mas também na inclusão da sobreposição da forma e volume moleculares (HAWKINS; SKILLMAN; NICHOLLS, 2007), e por esse diferencial em relação a outros métodos, selecionamos essa abordagem LBDD para complementar o trabalho. Os modelos *shape-based* construídos usando o programa ROCS tem se mostrado eficientes no processo de VS. Manoharan e colaboradores desenvolveram uma abordagem de triagem virtual baseada em fragmentos para identificação de inibidores de BACE1 (Beta-Secretase 1). Nesta VS foram usados modelos baseados na forma identificando quatro *hits* (MANOHARAN; GHOSHAL, 2018). A combinação dessas duas sobreposições (sobreposição de grupos e da forma e volume moleculares) já demonstrou ser eficiente na identificação de compostos ativos, com resultados comparáveis a métodos SBDD (MCGAUGHEY et al., 2007). Esta estratégia permitiu a seleção de compostos com boa sobreposição a inibidores já conhecidos da POPTc80, tanto em termos de forma (*shape*) quanto de características

moleculares específicas. Em nosso trabalho, a partir da disponibilidade de dados de inibição para a POPTc80, modelos *shape-based* foram gerados e validados com sucesso para serem usados durante o processo de VS.

Na etapa de ancoramento molecular, a validação dos modelos avaliou os resultados das funções de pontuação: *Docking score*, *Glide Emodel* e *Glide Energy*. *Docking score* é uma função de pontuação empírica que aproxima a energia livre de ligação do ligante. Ele tem muitos termos, incluindo contribuições de campo de força (eletrostático, van der Waals) e termos recompensando ou penalizando interações conhecidas por influenciar a ligação do ligante. Ele foi otimizado para precisão de encaixe, enriquecimento de banco de dados e previsão de afinidade de associação. O GlideScore deve ser usado para classificar poses de diferentes ligantes, por exemplo, em triagem virtual. Como simula uma energia livre de ligação, valores mais negativos representam ligantes mais apertados (FRIESNER et al., 2004, 2006).

Glide Emodel tem uma ponderação mais significativa dos componentes do campo de força (energias eletrostáticas e van der Waals), o que o torna adequado para comparar confôrmeros, mas muito menos para comparar espécies quimicamente distintas. Portanto, o Glide usa o *Emodel* para escolher a "melhor" pose de um ligante (seleção de pose) e, em seguida, classifica essas melhores poses umas contra as outras com o *GlideScore* (FRIESNER et al., 2004, 2006).

O Glide classifica seus resultados por *Docking score*. Isso significa que, se o usuário salvar várias poses por ligante, a classificação aparente das poses para um determinado ligante (pelo *Docking score*) não refletirá a classificação real do Glide usada para a seleção da pose. É sensato observar o valor das propriedades Emodel para determinar a "melhor" pose para um ligante. Desta forma, foram selecionados compostos melhor classificados pelo *Glide Emodel* com potencial de apresentarem boa afinidade ao sítio 2 da prolil oligopeptidase 80 de *Trypanosoma cruzi*. Apesar do potencial oferecido por este método, esforços crescentes têm sido direcionados para a melhoria dos algoritmos de *docking* e superação de suas limitações intrínsecas. Na verdade, grandes limitações do *docking* inclui a acurácia das funções de pontuação e a restrição da amostragem de conformações de macromoléculas e ligantes na predição de pose (BENTHAM SCIENCE PUBLISHER, 2006; ELOKELY; DOERKSEN, 2013; PANTSAR; POSO, 2018). Essas fraquezas inerentes são exacerbadas pela falta de experiência do usuário ao realizar etapas importantes, como a preparação de ligantes e proteínas, análise de enriquecimento, reclassificação de poses e outros. Essa

sequência de erros impacta negativamente as taxas de acerto nas triagens virtuais, levando à seleção aleatória e aquisição de compostos para validação experimental (NEVES et al., 2021).

A triagem virtual foi realizada utilizando tanto estratégias SBDD (*structure-based drug design*), como o ancoramento molecular, e também estratégias LBDD (*ligand-based drug design*). Na primeira etapa da triagem, os compostos da ChemBridge passaram pelo modelo *Shape-based*, que ao final desta etapa, foram selecionados apenas 1% do topo da lista, baseado nos melhores resultados do RefTverskyCombo. Logo após, os compostos selecionados passaram pelo modelo de ancoramento molecular, restando 67 moléculas. E por fim, esses compostos passaram por um filtro de agregadores e também de PAINS (*pan-assay interference compounds*), além de uma inspeção visual e análise de características físico-químicas, restando 15 *hits* virtuais ao final da triagem virtual.

O modo de interação desses *hits* se assemelha aos compostos inibidores da POP humana (PDB: 3DDU), sendo que o sítio ativo pode ser dividido em S1', S1, S2 e S3. O sítio S1' se caracteriza pelas interações com a Ser548, que faz parte da tríade catalítica da POPTc80. A região S1 é caracterizada por interações com o Trp589 e outros aminoácidos da região, formando uma região hidrofóbica para o fácil ajuste do substrato prolina ou dos anéis aromáticos inibidores. A região S2 é caracterizada por interações com o resíduo Arg633, sendo normalmente uma interação de hidrogênio. Por fim, a região S3 é formado por vários aminoácidos não polares, como a Cys255, criando uma grande cavidade hidrofóbica (KUMAR et al., 2017).

Os *hits* virtuais dessa pesquisa interagem de maneira diferente dos *hits* encontrados por Almeida e colaboradores (2016). Apesar dos ligantes apresentarem uma estrutura química semelhante aos *hits* desse trabalho, os aminoácidos envolvidos nas interações químicas são encontrados no sítio 1 descrito anteriormente nesse estudo. Entretanto, mesmo esses ligantes apresentarem atividade *in vitro*, os testes enzimáticos são incapazes de confirmar com quais resíduos esses compostos interagem. Em contrapartida, os inibidores co-cristalizados de estruturas homólogas à POPTc80 confirmam que os ligantes interagem no sítio 2. Diante do exposto, é possível assumir que o sítio 2 tem maior relevância e também melhores resultados estatísticos na validação do *docking* molecular.

Estes modelos validados também podem ser uma opção para a descoberta de possíveis inibidores de outras POPs, tais como, POP de *T. brucei* (POPTb) e POP de

Leishmania infantum (POPLi). Estes organismos possuem elevada semelhança de genes e podem ser alvos para busca de novos compostos inibidores de POP utilizando esses modelos validados de *docking* e *shape-based*. Uma outra possível aplicação para o modelo *shape-based* é a POP humana (PBD: EDDU), que é uma protease bastante relevante conforme já mencionado neste trabalho. Esse modelo *shape-based* específico para POP é uma inovação, possibilitando a redução do tempo de pesquisa e desenvolvimento de novos modelos.

CONCLUSÕES

No presente trabalho, um protocolo computacional foi desenvolvido para a triagem virtual de novos inibidores da POPTc80. Este protocolo, baseado em *docking* molecular e modelos *shape-based* apresentou elevada taxa de enriquecimento nos estudos de validação estatística, demonstrando desta forma sua capacidade de (i) identificar corretamente o sítio mais provável para ligação de inibidores de POPTc80; (ii) predizer a conformação potencialmente bioativa dos aminoácidos dos sítio ativo; e (iii) aumentar as taxas de recuperação de compostos inativos em triagens virtuais. Através desta abordagem, um conjunto de *hits* putativos foi priorizado e adquirido para avaliação experimental *in vitro* na POPTc80 e ensaios fenotípicos com *T. cruzi*. Uma vez apresentando atividade, estes poderão servir como bases moleculares para o planejamento inibidores de entrada celular para profilaxia e tratamento da doença de Chagas.

REFERÊNCIAS

- ALVAREZ, V. E. et al. Update on relevant trypanosome peptidases: Validated targets and future challenges. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, v. 1869, n. 2, p. 140577, fev. 2021.
- ASHKENAZY, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic acids research**, v. 44, n. W1, p. W344–W350, 8 jul. 2016.
- ASLETT, M. et al. TriTrypDB: A functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, n. SUPPL.1, 20 out. 2009.
- ATWOOD, J. A et al. The Trypanosoma cruzi proteome. **Science (New York, N.Y.)**, v. 309, n. 5733, p. 473–6, jul. 2005.
- BADRINARAYAN, P.; NARAHARI SASTRY, G. Virtual High Throughput Screening in New Lead Identification. **Combinatorial Chemistry & High Throughput Screening**, v. 14, n. 10, p. 840–860, 1 dez. 2011.
- BAELL, J. B.; HOLLOWAY, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. **Journal of Medicinal Chemistry**, v. 53, n. 7, p. 2719–2740, 8 abr. 2010.
- BAJORATH, J. **Computational chemistry in pharmaceutical research: At the crossroads**. **Journal of Computer-Aided Molecular Design**, jan. 2012.
- BAL, G. et al. Prolylisoxazoles: potent inhibitors of prolyl oligopeptidase with antitrypanosomal activity. **Bioorganic & Medicinal Chemistry Letters**, v. 13, n. 17, p. 2875–2878, set. 2003.
- BARRETT, A. J.; RAWLINGS, N. D.; O'BRIEN, E. A. The MEROPS Database as a Protease Information System. **Journal of Structural Biology**, v. 134, n. 2–3, p. 95–102, maio 2001.
- BARTSCH, S. M. et al. The economic value of identifying and treating Chagas disease patients earlier and the impact on Trypanosoma cruzi transmission. **PLOS Neglected Tropical Diseases**, v. 12, n. 11, p. e0006809, 5 nov. 2018.
- BASTOS, I. M. D. et al. Molecular, functional and structural properties of the prolyl oligopeptidase of Trypanosoma cruzi (POP Tc80), which is required for parasite entry into mammalian cells. **Biochemical Journal**, v. 388, n. 1, p. 29–38, 2005.

- BASTOS, I. M. D. et al. Prolyl oligopeptidase of *Trypanosoma brucei* hydrolyzes native collagen, peptide hormones and is active in the plasma of infected mice. **Microbes and Infection**, v. 12, n. 6, p. 457–466, jun. 2010.
- BENTHAM SCIENCE PUBLISHER, B. S. P. Docking and Scoring - Theoretically Easy, Practically Impossible? **Current Medicinal Chemistry**, v. 13, n. 25, p. 2995–3003, 1 out. 2006.
- BERN, C. Antitrypanosomal Therapy for Chronic Chagas' Disease. **New England Journal of Medicine**, v. 364, n. 26, p. 2527–2534, 30 jun. 2011.
- BERTHOLD, M. R. et al. KNIME - the Konstanz information miner. **ACM SIGKDD Explorations Newsletter**, v. 11, n. 1, p. 26–31, 16 nov. 2009.
- BIVONA, A. E. et al. *Trypanosoma cruzi* 80 kDa prolyl oligopeptidase (Tc80) as a novel immunogen for Chagas disease vaccine. **PLOS Neglected Tropical Diseases**, v. 12, n. 3, p. e0006384, 30 mar. 2018.
- BOGITSH, B. J. O. A. P. E. 2013. P. 85–113.; CARTER, C. E.; OELTMANN, T. N. Blood and Tissue Protozoa I: Hemoflagellates. In: BOGITSH, B. J.; CARTER, C. E.; OELTMANN, T. N. (Eds.). **Blood and Tissue Protozoa I: Hemoflagellates**. 4. ed. Oxford: Elsevier, 2013. p. 85–113.
- BORDOLI, L. et al. Protein structure homology modeling using SWISS-MODEL workspace. **Nature Protocols**, v. 4, n. 1, p. 1–13, 2009.
- BRAGA, R. et al. Virtual Screening Strategies in Medicinal Chemistry: The State of the Art and Current Challenges. **Current Topics in Medicinal Chemistry**, v. 14, n. 16, p. 1899–1912, 2014.
- BRAGA, R. C.; ANDRADE, C. H. Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good are They? **Current Topics in Medicinal Chemistry**, v. 13, n. 9, p. 1127–1138, 2013.
- BROOIJMANS, N.; KUNTZ, I. D. Molecular Recognition and Docking Algorithms. **Annual Review of Biophysics and Biomolecular Structure**, v. 32, n. 1, p. 335–373, jun. 2003.
- CALDAS, I. S.; SANTOS, E. G.; NOVAES, R. D. An evaluation of benznidazole as a Chagas disease therapeutic. **Expert Opinion on Pharmacotherapy**, v. 20, n. 15, p. 1797–1807, 13 out. 2019.
- CANÇADO, J. R. Criteria of Chagas disease cure. **Memórias do Instituto Oswaldo Cruz**, v. 94, n. suppl 1, p. 331–335, set. 1999.

- CAPRA, J. A.; SINGH, M. Predicting functionally important residues from sequence conservation. **Bioinformatics**, v. 23, n. 15, p. 1875–1882, ago. 2007.
- CASTRO, J. E. Z. Enfermedad de Chagas y otras tripanosomiasis. In: FLORES, M. A. B. (Ed.). **Parasitología médica**, 5e. New York, NY: McGraw-Hill Education, 2019.
- CAZZULO, J. Proteinases of Trypanosoma Cruzi: Potential Targets for the Chemotherapy of Chagas Disease. **Current Topics in Medicinal Chemistry**, v. 2, n. 11, p. 1261–1271, 1 nov. 2002.
- CAZZULO, J. J.; STOKA, V.; TURK V. Cruzipain, the major cysteine proteinase from the protozoan parasite Trypanosoma cruzi. **Biological Chemistry**, v. 378, n. 1, p. 1–10, 1997.
- CDC. **American Trypanosomiasis - Parasite Biology**.
- CHAGAS, C. Nova tripanozomíaze humana: estudos sobre a morfologia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., agente etiológico de nova entidade morbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, n. 2, p. 159–218, ago. 1909.
- CHAN, B. K. C. Data Analysis Using R Programming. In: [s.l: s.n.]. p. 47–122.
- CHAN, H. C. S. et al. Advancing Drug Discovery via Artificial Intelligence. **Trends in Pharmacological Sciences**, v. 40, n. 8, p. 592–604, ago. 2019.
- CHEN, Y.; POHLHAUS, D. T. In silico docking and scoring of fragments. **Drug Discovery Today: Technologies**, v. 7, n. 3, p. e149–e156, 1 set. 2010.
- CROWTHER, G. J. et al. Identification of attractive drug targets in neglected- disease pathogens using an in Silico approach. **PLoS Neglected Tropical Diseases**, v. 4, n. 8, ago. 2010.
- DAINA, A.; MICHELIN, O.; ZOETE, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. **Scientific Reports**, v. 7, n. 1, p. 42717, 3 maio 2017.
- DAVIES, D. R. The Structure and Function of the Aspartic Proteinases. **Annual Review of Biophysics and Biophysical Chemistry**, v. 19, n. 1, p. 189–215, jun. 1990.
- DE ALMEIDA, H. et al. Identification of novel Trypanosoma cruzi prolyl oligopeptidase inhibitors by structure-based virtual screening. **Journal of Computer-Aided Molecular Design**, v. 30, n. 12, p. 1165–1174, dez. 2016.
- DE TORANZO, E. G.; MASANA, M.; CASTRO, J. A. Administration of benznidazole, a chemotherapeutic agent against Chagas disease, to pregnant rats. Covalent binding of reactive metabolites to fetal and maternal proteins. **Archives**

internationales de pharmacodynamie et de therapie, v. 272, n. 1, p. 17–23, nov. 1984.

DEBNATH, B.; AL-MAWSAWI, L. Q.; NEAMATI, N. Are we living in the end of the blockbuster drug era? **Drug News & Perspectives**, v. 23, n. 10, p. 670–684, 2010.

DESJARLAIS, R. L. et al. Docking flexible ligands to macromolecular receptors by molecular shape. **Journal of Medicinal Chemistry**, v. 29, n. 11, p. 2149–2153, 1 nov. 1986.

DIMASI, J. A.; GRABOWSKI, H. G.; HANSEN, R. W. The Cost of Drug Development. **New England Journal of Medicine**, v. 372, n. 20, p. 1972–1972, maio 2015.

DNDI. Chagas Disease.

EKICI, Ö. D.; PAETZEL, M.; DALBEY, R. E. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. **Protein Science**, v. 17, n. 12, p. 2023–2037, 2 dez. 2008.

EKINS, S. et al. Exploiting machine learning for end-to-end drug discovery and development. **Nature Materials**, v. 18, n. 5, p. 435–441, 18 maio 2019.

ELDRIDGE, M. D. et al. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **Journal of Computer-Aided Molecular Design**, v. 11, n. 5, p. 425–445, 1997.

ELEBRING, T.; GILL, A.; PLOWRIGHT, A. T. **What is the most important approach in current drug discovery: Doing the right things or doing things right? Drug Discovery Today** Elsevier Ltd, , 2012.

ELOKELY, K. M.; DOERKSEN, R. J. Docking Challenge: Protein Sampling and Molecular Docking Performance. **Journal of Chemical Information and Modeling**, v. 53, n. 8, p. 1934–1945, 26 ago. 2013.

FABBRO, D. L. et al. Trypanocide treatment among adults with chronic Chagas disease living in Santa Fe city (Argentina), over a mean follow-up of 21 years: parasitological, serological and clinical evolution. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 40, n. 1, p. 1–10, fev. 2007.

FERREIRA, R. C.; FERREIRA, L. C. Mutagenicity of nifurtimox and benznidazole in the Salmonella/microsome assay. **Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas**, v. 19, n. 1, p. 19–25, 1986.

FIELD, M. C. et al. Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. **Nature reviews. Microbiology**, v. 15, n. 4, p. 217–231, fev. 2017.

FIOCRUZ. **ORGANIZAÇÃO ESTRUTURAL**.

FRANCO-PAREDES, C. American Trypanosomiasis. In: **Manson's Tropical Infectious Diseases**. [s.l.] Elsevier, 2014. p. 622-630.e1.

FRANKE DE CAZZULO, B. M. et al. Effects of proteinase inhibitors on the growth and differentiation of *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 124, n. 1, p. 81–86, nov. 1994.

FRIESNER, R. A. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1739–1749, 25 mar. 2004.

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, out. 2006.

FÜLÖP, V.; BÖCSKEI, Z.; POLGÁR, L. Prolyl Oligopeptidase: an unusual beta-propeller domain regulates proteolysis. **Cell**, v. 94, n. 2, p. 161–170, jul. 1998.

FÜLÖP, V.; SZELTNER, Z.; POLGÁR, L. Catalysis of serine oligopeptidases is controlled by a gating filter mechanism. **EMBO reports**, v. 1, n. 3, p. 277–281, set. 2000.

GASHAW, I. et al. **What makes a good drug target? Drug Discovery Today**, dez. 2011.

GIORDANO, R. et al. **MOLECULAR An acidic component from the surface of of the heterogeneous Trypanosoma cruzi is glycoprotein Tc-85 protein family a laminin bindingMolecular and Biochemical Parasitology**. [s.l: s.n.].

GIORDANO, R. et al. Cloning of a surface membrane glycoprotein specific for the infective form of *Trypanosoma cruzi* having adhesive properties to laminin. **Journal of Biological Chemistry**, v. 274, n. 6, p. 3461–3468, 5 fev. 1999.

GOLDBERG, A. L.; ROCK, K. L. Proteolysis, proteasomes and antigen presentation. **Nature**, v. 357, n. 6377, p. 375–379, jun. 1992.

GONZÁLEZ, J. et al. Proteasome activity is required for the stage-specific transformation of a protozoan parasite. **Journal of Experimental Medicine**, v. 184, n. 5, p. 1909–1918, 1 nov. 1996.

- GOODSELL, D. S. et al. Automated docking in crystallography: Analysis of the substrates of aconitase. **Proteins: Structure, Function, and Genetics**, v. 17, n. 1, p. 1–10, set. 1993.
- GOODSELL, D. S. et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery. **Protein Science**, v. 29, n. 1, p. 52–65, 29 jan. 2020.
- GORLA, N. B.; CASTRO, J. A. Micronucleus formation in bone marrow of mice treated with nifurtimox or benznidazole. **Toxicology Letters**, v. 25, n. 3, p. 259–263, jun. 1985.
- GRANDGENETT, P. M. et al. Differential expression of GP63 genes in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 110, n. 2, p. 409–415, out. 2000.
- GRANT, B. J.; SKJÆRVEN, L.; YAO, X. Q. The Bio3D packages for structural bioinformatics. **Protein Science**, v. 30, n. 1, p. 20–30, 1 jan. 2021.
- GRANT, J. A.; GALLARDO, M. A.; PICKUP, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. **Journal of Computational Chemistry**, v. 17, n. 14, p. 1653–1666, 1996.
- GRELLIER, P. et al. *Trypanosoma cruzi* Prolyl Oligopeptidase Tc80 Is Involved in Nonphagocytic Mammalian Cell Invasion by Trypomastigotes. **Journal of Biological Chemistry**, v. 276, n. 50, p. 47078–47086, 14 dez. 2001.
- GUEDES, I. A.; DE MAGALHÃES, C. S.; DARDENNE, L. E. Receptor–ligand molecular docking. **Biophysical Reviews**, v. 6, n. 1, p. 75–87, 21 mar. 2014.
- GUIDO, R.; OLIVA, G.; ANDRICOPULO, A. Virtual Screening and Its Integration with Modern Drug Design Technologies. **Current Medicinal Chemistry**, v. 15, n. 1, p. 37–46, 1 jan. 2008.
- HALL, B. S.; WILKINSON, S. R. Activation of Benznidazole by Trypanosomal Type I Nitroreductases Results in Glyoxal Formation. **Antimicrobial Agents and Chemotherapy**, v. 56, n. 1, p. 115–123, jan. 2012.
- HARDER, E. et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. **Journal of Chemical Theory and Computation**, v. 12, n. 1, p. 281–296, 12 jan. 2016.
- HART, T. N.; READ, R. J. A multiple-start Monte Carlo docking method. **Proteins: Structure, Function, and Genetics**, v. 13, n. 3, p. 206–222, jul. 1992.
- HAWKINS, P. C. D. et al. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge

structural database. **Journal of Chemical Information and Modeling**, v. 50, n. 4, p. 572–584, 26 abr. 2010.

HAWKINS, P. C. D.; SKILLMAN, A. G.; NICHOLLS, A. Comparison of shape-matching and docking as virtual screening tools. **Journal of Medicinal Chemistry**, v. 50, n. 1, p. 74–82, 1 jan. 2007.

HAY, M. et al. Clinical development success rates for investigational drugs. **Nature Biotechnology**, v. 32, n. 1, p. 40–51, jan. 2014.

HECKER, E. A. et al. Use of catalyst pharmacophore models for screening of large combinatorial libraries. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 5, p. 1204–1211, set. 2002.

HINTZE, B. J. et al. Molprobability's ultimate rotamer-library distributions for model validation. **Proteins: Structure, Function and Bioinformatics**, v. 84, n. 9, p. 1177–1189, mar. 2016.

HUA, S. et al. Purification and characterization of proteasomes from *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 78, n. 1–2, p. 33–46, jun. 1996.

HUANG, N.; SHOICHET, B. K.; IRWIN, J. J. Benchmarking Sets for Molecular Docking. **Journal of Medicinal Chemistry**, v. 49, n. 23, p. 6789–6801, 1 nov. 2006.

HUGHES, J. et al. Principles of early drug discovery. **British Journal of Pharmacology**, v. 162, n. 6, p. 1239–1249, mar. 2011.

IRWIN, J. J. et al. An Aggregation Advisor for Ligand Discovery. **Journal of Medicinal Chemistry**, v. 58, n. 17, p. 7076–87, set. 2015.

JA, I. MEga-trials for blockbusters. **JAMA**, v. 309, n. 3, p. 239–240, jan. 2013.

JACOBSSON, M. et al. Improving Structure-Based Virtual Screening by Multivariate Analysis of Scoring Data. **Journal of Medicinal Chemistry**, v. 46, n. 26, p. 5781–5789, 18 dez. 2003.

JAKALIAN, A.; JACK, D. B.; BAYLY, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. **Journal of Computational Chemistry**, v. 23, n. 16, p. 1623–1641, 18 out. 2002.

JOYEAU, R. et al. Synthesis and activity of pyrrolidinyl- and thiazolidinyl-dipeptide derivatives as inhibitors of the Tc80 prolyl oligopeptidase from *Trypanosoma cruzi*. **European Journal of Medicinal Chemistry**, v. 35, n. 2, p. 257–266, fev. 2000.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 26 ago. 2021.

- KAPETANOVIC, I. M. Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. **Chemico-Biological Interactions**, v. 171, n. 2, p. 165–176, 30 jan. 2008.
- KASZUBA, K. et al. Molecular dynamics, crystallography and mutagenesis studies on the substrate gating mechanism of prolyl oligopeptidase. **Biochimie**, v. 94, n. 6, p. 1398–1411, jun. 2012.
- KEARNES, S.; PANDE, V. ROCS-derived features for virtual screening. **Journal of Computer-Aided Molecular Design**, v. 30, n. 8, p. 609–617, 8 ago. 2016.
- KEENAN, M.; CHAPLIN, J. H. A New Era for Chagas Disease Drug Discovery? In: [s.l: s.n.]. p. 185–230.
- KHAMIS, M. A.; GOMAA, W.; AHMED, W. F. Machine learning in computational docking. **Artificial Intelligence in Medicine**, v. 63, n. 3, p. 135–152, mar. 2015.
- KIM, S. et al. PubChem 2019 update: improved access to chemical data. **Nucleic Acids Research**, v. 47, n. D1, p. D1102–D1109, 8 jan. 2019.
- KIRCHHOFF, L. V. Epidemiology of American Trypanosomiasis (Chagas Disease). In: [s.l: s.n.]. p. 1–18.
- 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, nov. 2004.
- KLEMBA, M.; GOLDBERG, D. E. Biological Roles of Proteases in Parasitic Protozoa. **Annual Review of Biochemistry**, v. 71, n. 1, p. 275–305, jun. 2002.
- KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. W1, jul. 2012.
- KOZAKOV, D. et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. **Nature Protocols**, v. 10, n. 5, p. 733–755, 6 maio 2015.
- KUMAR, R. et al. New compounds identified through in silico approaches reduce the α -synuclein expression by inhibiting prolyl oligopeptidase in vitro. **Scientific Reports**, v. 7, n. 1, p. 10827, 7 dez. 2017.
- KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions. **Journal of Molecular Biology**, v. 161, n. 2, p. 269–288, out. 1982.
- LAWANDI, J. et al. Inhibitors of Prolyl Oligopeptidases for the Therapy of Human Diseases: Defining Diseases and Inhibitors. **Journal of Medicinal Chemistry**, v. 53, n. 9, p. 3423–3438, 13 maio 2010.

- LEACH, A. R.; KUNTZ, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. **Journal of Computational Chemistry**, v. 13, n. 6, p. 730–748, jul. 1992.
- LEELANANDA, S. P.; LINDERT, S. Computational methods in drug discovery. **Beilstein Journal of Organic Chemistry**, v. 12, p. 2694–2718, 12 dez. 2016.
- LEIPART, V. et al. Structure prediction of honey bee vitellogenin: a multi-domain protein important for insect immunity. **FEBS Open Bio**, v. 12, n. 1, p. 51–70, 31 jan. 2022.
- LESCURE, F.-X. et al. Chagas disease: changes in knowledge and management. **The Lancet Infectious Diseases**, v. 10, n. 8, p. 556–570, ago. 2010.
- LI, H. et al. DynOmics: Dynamics of structural proteome and beyond. **Nucleic Acids Research**, v. 45, n. W1, p. W374–W380, 3 jul. 2017.
- LIDANI, K. C. F. et al. Chagas Disease: From Discovery to a Worldwide Health Problem. **Frontiers in Public Health**, v. 7, 2 jul. 2019.
- LIMA, A. N. et al. Use of machine learning approaches for novel drug discovery. **Expert Opinion on Drug Discovery**, v. 11, n. 3, p. 225–239, mar. 2016.
- LOMBARDINO, J. G.; LOWE, J. A. **The role of the medicinal chemist in drug discovery - Then and now**. **Nature Reviews Drug Discovery**, out. 2004.
- LOWNDES, C. M. et al. Heterogeneity of metalloprotease expression in *Trypanosoma cruzi*. **Parasitology**, v. 112, n. 4, p. 393–399, 6 abr. 1996.
- MA, L. et al. An evolutionary analysis of trypanosomatid GP63 proteases. **Parasitology Research**, v. 109, n. 4, p. 1075–1084, 19 out. 2011.
- MAES, M. et al. Alterations in plasma prolyl endopeptidase activity in depression, mania, and schizophrenia: effects of antidepressants, mood stabilizers, and antipsychotic drugs. **Psychiatry Research**, v. 58, n. 3, p. 217–225, out. 1995.
- MAES, M. EFFECTS OF PSYCHOLOGICAL STRESS ON SERUM PROLYL ENDOPEPTIDASE AND DIPEPTIDYL PEPTIDASE IV ACTIVITY IN HUMANS: HIGHER SERUM PROLYL ENDOPEPTIDASE ACTIVITY IS RELATED TO STRESS-INDUCED ANXIETY. **Psychoneuroendocrinology**, v. 23, n. 5, p. 485–495, jul. 1998.
- MAES, M. et al. Lower serum activity of prolyl endopeptidase in anorexia and bulimia nervosa. **Psychoneuroendocrinology**, v. 26, n. 1, p. 17–26, jan. 2001.
- MANOHARAN, P.; GHOSHAL, N. Fragment-based virtual screening approach and molecular dynamics simulation studies for identification of BACE1 inhibitor leads.

Journal of Biomolecular Structure and Dynamics, v. 36, n. 7, p. 1878–1892, 19 maio 2018.

MARTINS SANTANA, J. et al. **A Trypanosoma cruzi-secreted 80 kDa proteinase with specificity for human collagen types I and IV***Biochem. J.* [s.l: s.n.].

MAYA, J. D. et al. Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. **Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology**, v. 146, n. 4, p. 601–620, abr. 2007.

MCGAUGHEY, G. B. et al. Comparison of topological, shape, and docking methods in virtual screening. **Journal of Chemical Information and Modeling**, v. 47, n. 4, p. 1504–1519, jul. 2007.

MCKERROW, J. H. et al. THE PROTEASES AND PATHOGENICITY OF PARASITIC PROTOZOA. **Annual Review of Microbiology**, v. 47, n. 1, p. 821–853, out. 1993.

MEIRELLES, M. N. L. et al. Inhibitors of the major cysteinyl proteinase (GP57/51) impair host cell invasion and arrest the intracellular development of *Trypanosoma cruzi* in vitro. **Molecular and Biochemical Parasitology**, v. 52, n. 2, p. 175–184, jun. 1992.

MENDEZ, D. et al. ChEMBL: towards direct deposition of bioassay data. **Nucleic Acids Research**, v. 47, n. D1, p. D930–D940, 8 jan. 2019.

MENG, X.-Y. et al. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. **Current Computer Aided-Drug Design**, v. 7, n. 2, p. 146–157, 1 jun. 2011.

MILLS, J. E. J.; DEAN, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. **Journal of Computer-Aided Molecular Design**, v. 10, n. 6, p. 607–622, dez. 1996.

MUEGGE, I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. In: **Virtual Screening: An Alternative or Complement to High Throughput Screening?** Dordrecht: Kluwer Academic Publishers, 2000. p. 99–114.

MÜLLER KRATZ, J. et al. Clinical and pharmacological profile of benznidazole for treatment of Chagas disease. **Expert Review of Clinical Pharmacology**, v. 11, n. 10, p. 943–957, 3 out. 2018.

MURI, E. M. F. Viral proteases: important targets of peptidomimetic compounds. **Química Nova**, v. 37, n. 2, 2014.

NEVES, B. J. et al. Best Practices for Docking-Based Virtual Screening. In: **Molecular Docking for Computer-Aided Drug Design**. [s.l.] Elsevier, 2021. p. 75–98.

NICOLAOU, K. C. Advancing the Drug Discovery and Development Process. **Angewandte Chemie International Edition**, v. 53, n. 35, p. 9128–9140, 25 ago. 2014.

NWAKA, S.; RIDLEY, ROBERT. G. Virtual drug discovery and development for neglected diseases through public-private partnerships. v. 2, 2003.

OLLIS, D. L. et al. The α/β hydrolase fold. “**Protein Engineering, Design and Selection**,” v. 5, n. 3, p. 197–211, 1992.

OPENEYE SCIENTIFIC SOFTWARE. **ROCS**.

ORTEGA-BARRÍA, E. Trypanosoma Species (Trypanosomiasis). In: LONG, S. S.; PICKERING, L. K.; PROBER, C. G. (Eds.). **Principles and Practice of Pediatric Infectious Diseases**. 3. ed. London: Churchill Livingstone, 2008. p. 1290–1296.

PANTSAR, T.; POSO, A. Binding Affinity via Docking: Fact and Fiction. **Molecules**, v. 23, n. 8, p. 1899, 30 jul. 2018.

PATTERSON, S.; WYLLIE, S. Nitro drugs for the treatment of trypanosomatid diseases: past, present, and future prospects. **Trends in Parasitology**, v. 30, n. 6, p. 289–298, jun. 2014.

PÉREZ-MOLINA, J. A. et al. Old and new challenges in Chagas disease. **The Lancet Infectious Diseases**, v. 15, n. 11, p. 1347–1356, nov. 2015.

PÉREZ-MOLINA, J. A.; MOLINA, I. Chagas disease. **The Lancet**, v. 391, n. 10115, p. 82–94, jan. 2018.

POLGÁR, L. The prolyl oligopeptidase family. **Cellular and Molecular Life Sciences CMLS**, v. 59, n. 2, p. 349–362, fev. 2002.

PORCEL, B. M. et al. Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas’ disease. **The American Journal of Tropical Medicine and Hygiene**, v. 59, n. 4, p. 526–529, 1 out. 1998.

POSNER, B. A.; XI, H.; MILLS, J. E. J. Enhanced HTS Hit Selection via a Local Hit Rate Analysis. **Journal of Chemical Information and Modeling**, v. 49, n. 10, p. 2202–2210, 26 out. 2009.

- PUPKO, T. et al. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. **Bioinformatics**, v. 18, n. Suppl 1, p. S71–S77, 1 jul. 2002.
- RAJÃO, M. A. et al. Unveiling Benznidazole's mechanism of action through overexpression of DNA repair proteins in *Trypanosoma cruzi*. **Environmental and Molecular Mutagenesis**, v. 55, n. 4, p. 309–321, maio 2014.
- RASHID, U. et al. Structure based virtual screening-driven identification of monastrol as a potent urease inhibitor. **Journal of Molecular Graphics and Modelling**, v. 43, p. 47–57, jun. 2013.
- RAWLINGS, N. D.; BARRETT, A. J. [7] Families of aspartic peptidases, and those of unknown catalytic mechanism. In: [s.l: s.n.]. p. 105–120.
- RINGNÉR, M. What is principal component analysis? **Nature Biotechnology**, v. 26, n. 3, p. 303–304, mar. 2008.
- ROBERTS, L. S.; JAVONY, J. J. Kinetoplasta: Trypanosomes and Their Kin. In: ROBERTS, L. S.; JAVONY, J. J. (Eds.). **Foundations of Parasitology**. 8. ed. New York: McGraw-Hill, 2009. p. 61–88.
- ROSENTHAL, P. J. Proteases of Protozoan Parasites. In: [s.l: s.n.]. p. 105–159.
- SAJID, M.; MCKERROW, J. H. Cysteine proteases of parasitic organisms☆. **Molecular and Biochemical Parasitology**, v. 120, n. 1, p. 1–21, mar. 2002.
- SALES JUNIOR, P. A. et al. Experimental and Clinical Treatment of Chagas Disease: A Review. **The American Journal of Tropical Medicine and Hygiene**, v. 97, n. 5, p. 1289–1303, 8 nov. 2017.
- SCHNEIDER, G. Virtual screening: an endless staircase? **Nature Reviews Drug Discovery**, v. 9, n. 4, p. 273–276, abr. 2010.
- SCHRÖDINGER. Glide User Manual Glide 5.5 User Manual Glide User Manual. p. 118, 2009.
- SCHWEDE, T. Protein Modeling: What Happened to the “Protein Structure Gap”? **Structure**, v. 21, n. 9, p. 1531–1540, set. 2013.
- SCIOR, T. et al. Recognizing Pitfalls in Virtual Screening: A Critical Review. **Journal of Chemical Information and Modeling**, v. 52, n. 4, p. 867–881, 23 abr. 2012.
- SGAMBATTI DE ANDRADE, A. L. S. et al. Randomised trial of efficacy of benznidazole in treatment of early *Trypanosoma cruzi* infection. **The Lancet**, v. 348, n. 9039, p. 1407–1413, nov. 1996.

SHAN, L.; MATHEWS, I. I.; KHOSLA, C. **Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity**PNAS March. [s.l: s.n.]. Disponível em: <www.pdb.org>.

SHELLEY, J. C. et al. Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. **Journal of Computer-Aided Molecular Design**, v. 21, n. 12, p. 681–691, 27 dez. 2007.

SHEN, C. et al. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. **WIREs Computational Molecular Science**, v. 10, n. 1, 27 jan. 2020.

SHOICHET, B. K. Virtual screening of chemical libraries. **Nature**, v. 432, n. 7019, p. 862–865, 15 dez. 2004.

SILVA, C. L. **Estudo funcional da prolil oligopeptidase de Trypanosoma cruzie Leishmania infantum: nocaute gênico, inibição enzimática e processos infectivos**. [s.l: s.n.].

SIQUEIRA-NETO, J. L. et al. Cysteine proteases in protozoan parasites. **PLOS Neglected Tropical Diseases**, v. 12, n. 8, p. e0006512, 23 ago. 2018.

ŚLEDŹ, P.; CAFLISCH, A. Protein structure-based drug design: from docking to molecular dynamics. **Current Opinion in Structural Biology**, v. 48, p. 93–102, 2018.

SONEGO, P.; KOCSOR, A.; PONGOR, S. ROC analysis: Applications to the classification of biological sequences and 3D structures. **Briefings in Bioinformatics**, v. 9, n. 3, p. 198–209, 2008.

STERLING, T.; IRWIN, J. J. ZINC 15 – Ligand Discovery for Everyone. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2324–2337, 23 nov. 2015.

SUZEK, B. E. et al. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. **Bioinformatics**, v. 31, n. 6, p. 926–932, 15 mar. 2015.

SWANN, S. L. et al. A Unified, Probabilistic Framework for Structure- and Ligand-Based Virtual Screening. **Journal of Medicinal Chemistry**, v. 54, n. 5, p. 1223–1232, 10 mar. 2011.

TALELE, T. T.; KHEDKAR, S. A.; RIGBY, A. C. **Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic**Current Topics in Medicinal Chemistry. [s.l: s.n.].

TEIXEIRA, A. R. L.; CALIXTO, M. A.; TEIXEIRA, M. L. Chagas' disease: carcinogenic activity of the antitrypanosomal nitroarenes in mice. **Mutation**

Research/Fundamental and Molecular Mechanisms of Mutagenesis, v. 305, n. 2, p. 189–196, mar. 1994.

TEIXEIRA, D. E. et al. Interactive Multimedia to Teach the Life Cycle of *Trypanosoma cruzi*, the Causative Agent of Chagas Disease. **PLoS Neglected Tropical Diseases**, v. 6, n. 8, 28 ago. 2012.

TORRES, P. H. M. et al. Key Topics in Molecular Docking for Drug Design. **International Journal of Molecular Sciences**, v. 20, n. 18, p. 4574, set. 2019.

TROCHINE, A. et al. Benznidazole Biotransformation and Multiple Targets in *Trypanosoma cruzi* Revealed by Metabolomics. **PLoS Neglected Tropical Diseases**, v. 8, n. 5, p. e2844, 22 maio 2014.

TRUCHON, J. F.; BAYLY, C. I. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. **Journal of Chemical Information and Modeling**, v. 47, n. 2, p. 488–508, mar. 2007.

VENDEVILLE, S. et al. Identification of Inhibitors of an 80kDa Protease from *Trypanosoma cruzi* through the Screening of a Combinatorial Peptide Library. **CHEMICAL & PHARMACEUTICAL BULLETIN**, v. 47, n. 2, p. 194–198, 1999a.

VENDEVILLE, S. et al. Automated parallel synthesis of a tetrahydroisoquinolin-based library: Potential prolyl endopeptidase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 9, n. 3, p. 437–442, fev. 1999b.

VERLI, H. **Bioinformática da Biologia à Flexibilidade Molecular**. [s.l.: s.n.]. v. 53

VIOTTI, R. et al. Long-Term Cardiac Outcomes of Treating Chronic Chagas Disease with Benznidazole versus No Treatment. **Annals of Internal Medicine**, v. 144, n. 10, p. 724, 16 maio 2006.

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.

WERMUTH, C. et al. Glossary of terms used in medicinal chemistry. **Pure and Applied Chemistry**, v. 70, n. 5, p. 1129–1143, 1998.

WILKESMAN, J. Cysteine Protease Zymography: Brief Review. In: [s.l.: s.n.]. p. 25–31.

WILKINSON, S. R.; KELLY, J. M. Trypanocidal drugs: mechanisms, resistance and new targets. **Expert Reviews in Molecular Medicine**, v. 11, p. e31, 29 out. 2009.

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

WOLBER, G.; LANGER, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 160–169, 1 jan. 2005.

WOUTERS, O. J.; MCKEE, M.; LUYTEN, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. **JAMA**, v. 323, n. 9, p. 844, 3 mar. 2020.

XU, P.; HUANG, M. Small Peptides as Modulators of Serine Proteases. **Current Medicinal Chemistry**, v. 27, n. 22, p. 3686–3705, 30 jun. 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.

YOUNG, D. C. COMPUTATIONAL DRUG DESIGN. In: **Computational Drug Design**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2009. p. i–xxxvi.

YUN, O. et al. Feasibility, Drug Safety, and Effectiveness of Etiological Treatment Programs for Chagas Disease in Honduras, Guatemala, and Bolivia: 10-Year Experience of Médecins Sans Frontières. **PLoS Neglected Tropical Diseases**, v. 3, n. 7, p. e488, 7 jul. 2009.