



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS DA SAÚDE – CCS
PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUTOS NATURAIS E SINTÉTICOS
BIOATIVOS – PgPNSB**

IGOR MIKAEL ALVES DE ARAUJO

**TRIAGEM VIRTUAL DE ALCALOIDES ISOLADOS DA FAMÍLIA
EUPHORBIACEAE FRENTE A DOENÇA DE CHAGAS**

**João Pessoa – PB
Agosto de 2023**

IGOR MIKAEL ALVES DE ARAUJO

**TRIAGEM VIRTUAL DE ALCALOIDES ISOLADOS DA FAMÍLIA
EUPHORBIACEAE FRENTE A DOENÇA DE CHAGAS**

Dissertação de Mestrado do Programa de Pós graduação em Produtos Naturais e Sintéticos Bioativos da UFPB, como requisito para a obtenção do título de Mestre em Produtos Naturais e Sintéticos Bioativos na Área de concentração: Farmacoquímica.

Área de Concentração: Farmacoquímica.

Orientador: Prof. Dr. Marcus Tullius Scotti

**João Pessoa – PB
Agosto de 2023**

Catálogo na publicação
Seção de Catalogação e Classificação

A663t Araujo, Igor Mikael Alves de.

Triagem virtual de alcaloides isolados da família
Euphorbiaceae frente a Doença de Chagas / Igor Mikael
Alves de Araujo. - João Pessoa, 2023.

132 f. : il.

Orientação: Marcus Tullius Scotti.
Dissertação (Mestrado) - UFPB/CCS.

1. Doença de Chagas. 2. Triagem virtual. 3. Produtos
naturais - Banco de dados. 4. Alcaloides. 5. Moléculas
multitarget. I. Scotti, Marcus Tullius. II. Título.

UFPB/BC

CDU 616.937(043)




PgPNSB

Pós Graduação em Produtos Naturais
e Sintéticos Bioativos


Ata da 483^a (quadringentésima octogésima terceira) Dissertação de Mestrado do(a) aluno(a) do Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos **Igor Mikael Alves de Araújo**, candidato(a) ao Título de “Mestre” em Produtos Naturais e Sintéticos Bioativos na área de concentração Farmacoquímica.

Às nove horas e trinta minutos (09h30) do dia trinta e um de agosto do ano de dois mil e vinte e três (31/08/2023), em ambiente virtual de videoconferência através do aplicativo Google Meet, link: <https://meet.google.com/dyr-wvuq-cng>, reuniram-se em caráter de Solenidade Pública os membros da Comissão designada para examinar o(a) discente **Igor Mikael Alves de Araújo**, candidato(a) ao Título de “MESTRE” em Produtos Naturais e Sintéticos Bioativos na área de concentração Farmacoquímica. Foram componentes da Comissão Examinadora os pesquisadores Jefferson Almeida Rocha, Ph.D em Biotecnologia, Sócrates Golzio dos Santos, Ph.D em Farmacoquímica, e Marcus Tullius Scotti, Ph.D em Química Orgânica. Sendo o primeiro, integrante do corpo docente da Universidade Federal do Maranhão; e os demais, integrantes do corpo docente da Universidade Federal da Paraíba. Dando início aos trabalhos, o(a) Presidente da Comissão, professor(a) Marcus Tullius Scotti, após declarar os objetivos da reunião, apresentou o(a) candidato(a) **Igor Mikael Alves de Araújo**, a quem concedeu a palavra para que dissertasse oral e sucintamente sobre o tema apresentado e intitulado “Triagem virtual de alcaloides isolados da família Euphorbiaceae frente a Doença de Chagas”. Após discorrer sobre o referido tema durante cerca de quarenta minutos, o(a) candidato(a) foi arguido(a) pelos examinadores na forma regimental. Em seguida, passou a comissão, em caráter secreto, a proceder à avaliação e julgamento do trabalho, concluindo por atribuir-lhe o conceito APROVADO(A). Em face da aprovação, declarou o(a) Presidente achar-se o(a) examinado(a) **Igor Mikael Alves de Araújo** legalmente habilitado(a) a receber o Título de “MESTRE” em Produtos Naturais e Sintéticos Bioativos, na área de concentração Farmacoquímica, cabendo a Universidade Federal da Paraíba, providências, como de direito, a expedição do Diploma que o(a) mesmo(a) faz jus. Nada mais havendo a tratar, foi lavrada a presente ata que é abaixo assinada pelos membros da Comissão e pelo(a) discente.

Prof. Dr. Marcus Tullius Scotti
(Orientador)

Documento assinado digitalmente
 **MARCUS TULLIUS SCOTTI**
Data: 18/09/2023 18:15:45-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Jefferson Almeida
Rocha (Examinador)

Documento assinado digitalmente
 **JEFFERSON ALMEIDA ROCHA**
Data: 14/09/2023 17:30:30-0300
Verifique em <https://validar.iti.gov.br>





PgPNSB

Pós Graduação em Produtos Naturais
e Sintéticos Bioativos

Prof. Dr. Sócrates Golzio dos
Santos (Examinador)

Documento assinado digitalmente
gov.br SOCRATES GOLZIO DOS SANTOS
Data: 12/09/2023 17:42:00-0300
Verifique em <https://validar.iti.gov.br>

Igor Mikael Alves de Araújo
(Discente)

Documento assinado digitalmente
gov.br IGOR MIKAEL ALVES DE ARAUJO
Data: 14/09/2023 18:08:50-0300
Verifique em <https://validar.iti.gov.br>



DEDICATÓRIA

Dedico este trabalho a minha mãe Benedita Alves por todo amor e compreensão, ao meu pai Janildo Oliveira, pelo amor e apoio, aos meus irmãos, Hugo Vinyus e Izadora Alves, pelo companheirismo e por toda força.

AGRADECIMENTOS

Agradeço primeiramente a Deus por cada conquista, por me encorajar a vencer todos os desafios e por toda força para o alcance dos meus objetivos. Aos meus pais Benedita Alves e Janildo Oliveira por sempre me apoiarem em todos os meus sonhos. Mesmo longe me fortaleceram com seu amor, apoio e auxílio nesta jornada. Acredito que sem o apoio deles seria muito difícil vencer este desafio.

Aos meus irmãos, Hugo e Izadora, por todo auxílio, amor e cuidados. Ao meu tio Jailson, que sempre me ajudou e acreditou na minha vitória. As minhas tias Isabel e Eliene pela ajuda e incentivos mesmo à distância. Aos meus avós Olga, Benedito e Maria, por todo incentivo, amor e cuidado. Agradeço também à minha companheira Camila, que esteve comigo nessa caminhada, compartilhando momentos de alegria e de tristeza, e que mesmo com todos os desafios não me deixou abater pelas adversidades da vida que por vezes se apresentavam. Sou muito grato a ela pela paciência e pelo carinho em todo esse tempo.

Ao meu orientador Prof. Dr. Marcus Tullius Scotti, pela oportunidade, apoio, confiança, atenção e tempo dedicado na elaboração deste trabalho.

A Professora Dr^a. Luciana Scotti pela sua disponibilidade, atenção e gentileza em todos os momentos e por ter me orientado no início do mestrado.

A banca avaliadora deste trabalho, composta pelos Professores: Dr. Sócrates Golzio dos Santos e Dr Jefferson Almeida Rocha, merece nosso sincero agradecimento por ter aceitado o convite e por suas valiosas contribuições.

A Universidade Federal da Paraíba (UFPB), ao IPeFarm e à Coordenação do Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos. Aos demais Docentes do Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos pela oferta das disciplinas e por todo o conhecimento transmitido.

A minha amiga e companheira do Laboratório de Quimioinformática, doutoranda Natalia Ferreira pela ajuda e apoio constantes ao longo desta pesquisa, além da amizade que construímos durante esse trajeto. A você, expresso toda a minha gratidão e amizade!

A todos os colegas que compõe o laboratório de Quimioinformática.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida e por todo o apoio financeiro.

Aos meus amigos que o curso de Química me proporcionou, gostaria de expressar minha gratidão por todo o apoio e incentivo, mesmo que à distância. Vocês estiveram presentes

de maneira significativa durante essa jornada. Em particular, gostaria de mencionar Zelma, Carol, Cristiano e Dhemerson, que compartilharam essa etapa comigo.

Por fim, agradeço a todos que direta ou indiretamente fizeram parte da minha formação!

EPÍGRAFE

“Que a ciência possa gerar cada vez mais dúvidas e incertezas, para que a cada trabalho, tenhamos novas perguntas e que a cada pergunta possamos ter nova resposta.”

André Lima

RESUMO

A Doença de Chagas é uma infecção causada pelo parasita *Trypanosoma cruzi* e é endêmica em vários países da América Latina. O tratamento para a doença de Chagas é limitado a alguns medicamentos, além disso apresentam limitações, incluindo efeitos colaterais significativos e eficácia variável nas diferentes fases da doença. Devido isso há uma necessidade urgente de novos medicamentos para tratar as formas evolutivas do parasita. Produtos naturais, como os derivados de plantas da família Euphorbiaceae, são considerados uma potencial fonte de novos tratamentos. Essas plantas são amplamente utilizadas na medicina tradicional e contêm diversos metabólitos secundários. O objetivo principal foi a identificação compostos com potencial atividade biológica contra a doença de Chagas através de uma triagem virtual. O foco foram alcaloides isolados da família Euphorbiaceae. Dessa forma, um banco de dados de alcaloides isolados da família Euphorbiaceae foi construído, visando à elaboração de um perfil químico. Um modelo de predição foi elaborado no KNIME. Para isso, o conjunto de dados foi obtido da base de dados ChEMBL, e os compostos foram classificados de acordo com os valores de pIC50 e o cálculo de descritores foi realizado através do software Volsurf. A execução da docagem molecular foi conduzida no software Molegro Virtual Docker, juntamente com a análise das interações para quatro proteínas obtidas no Protein Data Bank. Além disso, as funções de pontuação Moldock score, PLANTS score e Rerank score foram utilizadas, e as energias de ligação foram consensualmente avaliadas. O modelo preditivo desenvolvimento classificou moléculas com probabilidade acima de 70% para a forma tripromastigota de *T. cruzi*, resultando na identificação de 21 moléculas com potencial atividade. As análises de docking molecular foram positivas, indicando interações dos compostos selecionados com as enzimas alvo, evidenciadas por energias negativas. Quanto à absorção, os compostos demonstraram mais de 55% de absorção por via oral, com boa disponibilidade, geralmente com apenas uma violação nas regras. Na análise de toxicidade, somente nove compostos exibiram indícios de toxicidade em um ou dois parâmetros. Após o cálculo dos valores de probabilidade combinada (ancoragem molecular e modelo de predição), a molécula 149 foi selecionada por apresentar um maior percentual em todas as enzimas, demonstrando assim um potencial *multitarget*. Os resultados das simulações de dinâmica molecular demonstram que o RMSD do complexo para o composto teste molécula 149 (Magnoflorina) permanece estável. Além disso, pode-se sugerir que por meio dos cálculos de RMSF e energia de interação, o composto teste Magnoflorina interage com este alvo, possibilitando interação, flexibilidade e estabilidade.

Palavras-chave: Doença de Chagas; Triagem Virtual; Banco de Dados de Produtos Naturais; Alcaloides; moléculas *multitarget*.

ABSTRACT

Chagas disease is an infection caused by the *Trypanosoma cruzi* parasite and is endemic in several Latin American countries. Treatment for Chagas disease is limited to a few medications, which also have limitations, including significant side effects and variable effectiveness at different stages of the disease. Because of this there is an urgent need for new drugs to treat the evolving forms of the parasite. Natural products, such as those derived from plants of the Euphorbiaceae family, are considered a potential source of new treatments. These plants are widely used in traditional medicine and contain several secondary metabolites. The main objective was to identify compounds with potential biological activity against Chagas disease through virtual screening. The focus was isolated alkaloids from the Euphorbiaceae family. Thus, a database of alkaloids isolated from the Euphorbiaceae family was built, aiming at the elaboration of a chemical profile. A prediction model was built in KNIME. For this, the data set was obtained from the ChEMBL database, and the compounds were classified according to the pIC₅₀ values and the calculation of descriptors was performed using the Volsurf software. Molecular docking was carried out using the Molegro Virtual Docker software, along with the analysis of interactions for four proteins obtained from the Protein Data Bank. In addition, the Moldock score, PLANTS score and Rerank score functions were used, and the binding energies were consensually evaluated. The developed predictive model classified molecules with a probability above 70% for the *T. cruzi* tripromastigotes, resulting in the identification of 21 molecules with potential activity. Molecular docking analyzes were positive, indicating interactions of the selected compounds with the target enzymes, evidenced by negative energies. As for absorption, the compounds have demonstrated greater than 55% absorption orally, with good availability, usually with only one rule violation. In the toxicity analysis, only nine compounds showed signs of toxicity in one or two parameters. After calculating the combined probability values (molecular anchorage and prediction model), molecule 149 was selected because it presented a higher percentage in all enzymes, thus demonstrating a multitarget potential. The results of molecular dynamics simulations demonstrate that the RMSD of the complex for the test compound molecule 149 (Magnoflorina) remains stable. Furthermore, it can be suggested that through calculations of RMSF and interaction energy, the test compound Magnoflorina interacts with this target, enabling interaction, flexibility and stability.

Keywords: Chagas disease; Virtual Screening; Natural Products Database; alkaloids; *multitarget* molecules.

LISTA DE ABREVIATURAS, SIGLAS E FÓRMULAS

2D – Two dimensional (Bidimensional)
3D – Three-dimensional (Tridimensional)
A.M - Ancoragem Molecular
ABS – Absorção
ABS- Absorção
ADMET – Absorção, Distribuição, Metabolismo, Excreção e Toxicidade
ATB– Automated Topology Builder (Construtor de topologia automatizado)
AUC - Area Under the Curve (Área sob a curva)
BBB: Barreira Hematoencefálica; P-Glicoproteína
BNZ - Benznidazol
CPCA – Consensus Principal Component Analysis (Consenso das análises de componentes principais)
CYP: Modelos de Metabolismo de Citocromo
DA – Applicability Domain (Domínio de Aplicabilidade)
DC- Doença de Chagas
DHFR - Dihydrofolate reductase-thymidylate synthase
DM - Dinâmica Molecular
DM- Dinamica Molecular
DTU- *Unidades Discretas de Tipagem*
ESR: Efeito no Sistema Reprodutor
FN – False Negative (Falso Negativo)
FN- Falso negativo
FN- Falso negativo
FP – False Positivo (Falso Positivo)
FP- Falso positivo
GI: Absorção Gastrointestinal
GRIND GRid- INdependent Descriptors- Descritores independentes de alinhamento
GRIND- GRid-INdependent Descriptors (Descritores independentes de alinhamento)
HTS - High-throughput screening
IC50 – Inhibition Concentration 50% (Concentração Inibitória 50%)
IgG- Imunoglobulina G
INF- γ - Interferon-gama

IRR- Irritabilidade do tecido
IRR: Irritabilidade no Tecido
LBVS -Ligand Based Virtual Screening (Triagem Virtual Baseada em Ligante)
LCR - Líquido cefalorraquidiano
LogP: coeficiente de partição
MCC - Coeficiente de Correlação de Matthews
MIFs - Molecular Interaction Fields (Campos de Interação Molecular)
MS – MolDock Score
MUT: Mutagenicidade
MVD – Molegro Virtual Docker
MVT – Número Constante de Partículas, Volume e Temperatura)
NK- Natural Killer(Assassino natural)
OMS- Organização Mundial da Saúde
PCA – Principal Component Analysis (Análises de Componentes Principais)
PCR- Reação em Cadeia da Polimerase
PDB – Protein Data Bank (banco de dados de proteínas)
PLS- Partial Least Squares (Mínimos Quadrados Parciais)
QSAR – Quantitative-Structure Activity Relationship (Relação Quantitativa Estrutura-Atividade)
RF- Random Forest (floresta aleatória)
RMSD – Root Mean Square Deviation (Raiz do Desvio Quadrático Médio)
RMSF– Residual Fluctuations (Flutuações Residuais)
ROC – Receiver Operating Characteristic (Característica de Operação do Receptor)
SDF- Spatial Data File (Arquivo de Dados Espaciais)
SDF- Structure Data Format (Formato de dados da estrutura)
SFM - Sistema Fagocítico Mononuclear
SMILES- Simplified Molecular Input Line Entry Specification (Especificação Simplificada de Entrada de Linha de Entrada Molecular)
SNC- Sistema Nervoso Central
SPC- Simple Point Charge (Carga Pontual Simples)
TBVS- Target-Based Virtual Screening (Triagem Virtual Baseada no Alvo)
TN- True Negative (Verdadeiro Negativo)
TNF- α - Fator de Necrose Tumoral- Alfa
TOX: Toxicida

TP – True Positive (Verdadeiro Positivo)

TP- True Positive (Verdadeiro Positivo)

TPSA- Área de Superfície Topológica Total

TUM: Tumorigenicidade

VN- Taxa de verdadeiros negativos do modelo

VOLSURF, - Volume and Surface (Volume e Superfície)

LISTA DE FIGURAS

Figura 1. Espécie do gênero Croton (<i>Croton heliotropiifolius</i>)	22
Figura 2. Espécie do gênero Euphorbia (<i>Euphorbia hirta</i>)	23
Figura 3. Parasita Trypanosoma cruzi	25
Figura 4. Ciclo de vida do <i>T. Cruzi</i>	27
Figura 5. Estrutura 2D do Nifurtimox (A) e Benznidazol (B)	31
Figura 6. Fluxo de trabalho (Workflow) empregado na construção do modelo e realização das análises estatísticas	50
Figura 7. Principais gêneros dos alcaloides de Euphorbiaceae presente no banco de dados .	54
Figura 8. Gráfico ROC com área sob a curva para o conjunto de teste do modelo de T cruzi obtido com “Random Forest ”	57
Figura 9. Compostos moleculares (arquitetura 2D) em estudo com maiores percentuais de probabilidade de atividade para a forma tripomastigota de T cruzi	58
Figura 10. Ligantes e as interações correspondentes no sítio ativo das enzimas em 3D (A) e 2D(B).....	64
Figura 11. Interações 3D (A) e 2D (B) de alcaloides com os melhores valores de energia de ancoragem frente a <i>T. cruzi</i>	72
Figura 12. Interações realizadas pelo composto Magnoflorine com os resíduos de aminoácidos correspondentes, proteínas 1GXF; 1MXF; 3HBB; 3LXS, em 3D (A) e 2D (B)	80
Figura 13. RMSD de átomos C α da proteína Diidrofolato redutase-timidilato sintase de Trypanosoma cruzi (PDB: 3HBB) (linha preta) e complexado ao composto de teste Magnoflorine (linha vermelha) e ao ligante PDB Trimetrexato (linha verde)..	88
Figura 14. RMSD dos átomos C α dos compostos. Composto de teste Magnoflorine (linha preta) e ligante PDB Trimetrexato (linha vermelha).....	89
Figura 15. RMSF de átomos. Alvo Diidrofolato redutase-timidilato sintase de Trypanosoma cruzi (PDB: 3HBB) (linha preta) complexado ao composto de teste Magnoflorine (linha vermelha) e ligante PDB Trimetrexato (linha verde).....	90

LISTA DE TABELAS

Tabela 1. Informações das enzimas alvos para T cruzi selecionadas, depositadas no banco de dados PDB e usadas para ancoragem	51
Tabela 2. Resultados dos parâmetros estatísticos de confiabilidade do modelo de predição com compostos químicos de atividade conhecida para forma tripomastigota utilizando os descritores VolSurf	56
Tabela 3. Valores referente as pontuações do MolDock para o ligante e RMSD	63
Tabela 4. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 1GXF	67
Tabela 5. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 1MXF	68
Tabela 6. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 3HBB	69
Tabela 7. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 3LXS	70
Tabela 8. Valores referente a pontuação do Moldock para o ligante e das moléculas que obtiveram menores energias e maiores valores de probabilidade	71
Tabela 9. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 1GXF	76
Tabela 10. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 1MXF	77
Tabela 11. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 3HBB	77
Tabela 12. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 3LXS	78
Tabela 13. Molécula com Potencial Ativo e multitarget	79
Tabela 14. Propriedades físico-químicas e farmacocinéticas do ADMET para os alcaloides	85
Tabela 15. Resultado das análises de toxicidade dos alcaloides em estudo	86
Tabela 16. Valores de energia de interação de Coulombic e Lennard-Jones	90

SUMÁRIO

1	INTRODUÇÃO	16
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	PRODUTOS NATURAIS	18
2.2	FAMILIA EUPHORBIACEAE.....	20
2.3	DOENÇA DE CHAGAS	24
2.4	ALCALOIDES.....	31
2.5	TRIAGEM VIRTUAL.....	33
2.6	PREDIÇÃO DE DADOS	35
2.7	DOCAGEM MOLECULAR	36
2.8	DINÂMICA MOLECULAR	40
3	OBJETIVOS	41
3.1	OBJETIVO GERAL	41
3.2	OBJETIVOS ESPECIFICOS.....	41
4	METODOLOGIA	41
4.1	BANCO DE DADOS	41
4.2	DESCRITORES VOLSURF	43
4.3	MODELO DE PREDIÇÃO.....	44
4.3.1	<i>KNIME Analytics</i>	44
4.3.2	<i>Random Forest</i>	45
4.3.3	<i>Wecka Predictor 3.7</i>	46
4.3.4	<i>Curva ROC</i>	47
4.4	DOCAGEM MOLECULAR	50
4.5	PREVISÃO DE PROPRIEDADES ADMET.....	51
4.6	SIMULAÇÕES DE DINÂMICA MOLECULAR.....	52
5	RESULTADOS E DISCUSSÃO	52
5.1	ANALISE DO BANCO DE DADOS	52
5.2	ANALISE DOS MODELOS DE <i>T CRUZI</i> NA FORMA <i>TRIPOMASTIGOSTA</i>	53
5.2.1	<i>Descritores VolSurf</i>	54
5.3	RESULTADOS DOCKING MOLECULAR.....	61
5.4	PREVISÃO DE PROPRIEDADES ADMET	81
5.5	RESULTADO DA DINÂMICA MOLECULAR	85
6	CONCLUSÃO	89
	REFERENCIAS	91
	ANEXOS	109

2 INTRODUÇÃO

Historicamente, as doenças negligenciadas afligem a população mundial, afetando principalmente a população mais pobre e vulnerável e contribuindo de forma significativa na perpetuação dos ciclos de pobreza, desigualdade e exclusão social [1]. O desenvolvimento de pesquisas inovadoras de novos fármacos contra essas doenças é urgentemente necessário, e as plantas medicinais podem desempenhar um papel importante, já que os metabolitos secundários produzidos pelas plantas possuem uma imensurável riqueza de produtos bioativos [4].

Evidências históricas apontam que, desde o surgimento da humanidade, o uso de plantas medicinais é uma técnica antiga, amplamente utilizada para fins terapêuticos no tratamento, cura e prevenção de enfermidade. As plantas possuem uma ampla e rica composição química. Dessa forma, atualmente, diversos estudos são realizados em busca de novos fármacos [2].

Dentre as investigações realizadas com produtos naturais, a família Euphorbiaceae é uma das famílias de plantas mais estudadas no campo das doenças negligenciadas, pois tem permitido o isolamento de uma grande variedade de substâncias que exibem diversas atividades biológicas, possibilitando a utilização como marcadores aplicáveis à taxonomia em domínios hierárquicos, localizados nos níveis mais baixos. Contudo, ao longo dos últimos anos, é crescente a diversidade de estudos relacionados a triagem virtual, com grande ênfase na atividade antiprotozoária [3].

Os produtos naturais causam grande impacto na humanidade, pois oferecem moléculas promissoras, que resultam na predominância da produção de infinitos metabolitos secundários com diversas atividades biológicas [4]; [118]. Diversos estudos já validaram o efeito de produtos naturais como potenciais fontes de seletivos agentes para o tratamento de doenças tropicais causados por protozoários e outros parasitas [119]; [120]. Vários constituintes químicos encontrados em estudos relacionam-se com os efeitos terapêuticos descritos para espécies de Euphorbiaceae, o que justifica sua utilização medicinal. Esse fato tem levado ao isolamento de uma grande variedade de substâncias [3].

Os avanços da química medicinal moderna têm possibilitado o envolvimento de métodos computacionais, como a modelagem molecular, que contribuem significativamente para o desenvolvimento de diversos estudos. A evolução desses estudos por meio dessa ferramenta, nos últimos anos, tem emergido perpetuamente na área de química medicinal, proporcionando a descoberta de novos fármacos cada vez mais seguros e eficazes, além de desenvolver novos tratamentos com menores custos e recursos financeiros [5]; [80]; [81].

A quimioinformática é uma ciência baseada em técnicas computacionais que auxiliam na descoberta de novos medicamentos, através do arranjo de bioativos inéditos, por intermédio dos diversos métodos e ferramentas *in silico*. Diversos estudos dessa área de concentração executado mediante utilização desses métodos, são desenvolvidas por pesquisadores com a finalidade de prever a atividade biológica de compostos promissores. Isso é feito a partir dos seus diversos parâmetros estruturais com possível atividade, baseada na estrutura molecular, com estudos biológicos já testado previamente *in vitro* ou *in vivo* [85]; [86].

Estudos *in silico* desenvolvido mediante ferramentas de química computacional abrangem o foco dos estudos para aplicações biológicas. Através do gerenciamento de dados informacionais, a ampliação da possibilidade de êxito nas pesquisas direcionadas a análise de dados advém da eficácia do diagnóstico e administração das informações catalogadas [87].

As técnicas de triagem virtual se consolidam como uma ferramenta computacional de grande relevância para a química medicinal, principalmente na descoberta de novos fármacos. Essa ferramenta possibilita na orientação e, conseqüentemente, a identificação de novos medicamentos por meio de extensas bibliotecas de moléculas químicas, objetivando a seleção das melhores moléculas por meio de diversas técnicas computacionais (*in silico*). Dessa forma, é possível realizar a identificação de estruturas com maior probabilidade de se ligar a um determinado alvo farmacológico, geralmente um receptor ou enzima [6].

Dentre os diversos metabolitos secundários isolados, destacam-se terpenóides; cumarinas, fenóis, lignanas, taninos hidrolisáveis, terpenos (monoterpenos e triterpenos), flavonoides (flavonóis e flavanonas), alcaloides, triterpenos e esteroides, diterpenos [7]; [8]; [9]; [10]. Na contemporaneidade, todas as classes que são representados por no mínimo algumas centenas, o que possibilita o estudo da fitoquímica comparada e associada a taxonomia (quimiotaxonomia). Dessa forma, é importante destacar que os produtos naturais, quando estudados associados a químicos sintéticos e biológicos, são promissores e com extensa potencialidade. Isso visa à descoberta de novas estruturas com grande potencial para a descoberta de candidatos a fármacos para doenças humanas [4].

Nesse contexto, a pesquisa foi executada através de ferramentas computacionais baseados na estrutura do ligante, como também baseado na estrutura do receptor, além de outros métodos de triagem virtual. O objetivo geral a identificação de novos alcaloides com potenciais atividade anti-chagásica, por meio de um banco de dados de alcaloides da família Euphorbiaceae.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Produtos Naturais

A utilização de produtos naturais ao longo dos séculos construiu a base no tratamento e cura de doenças, um processo que ano após ano permanece evoluindo e consolidando-se na indústria farmacêutica em vários países e regiões por todo o mundo [11]. Desde os primórdios, a medicina tradicional baseada na utilização de várias ervas, para o alívio e prevenção de enfermidades, fez parte historicamente da cultura de diversas sociedades, desde a antiguidade até a contemporaneidade. Os produtos de origem vegetal contribuíram amplamente no desenvolvimento do mercado farmacêutico e ainda permanecem atualmente desempenhando um papel importante e crucial nas descobertas significativas e promissoras de novos fármacos [4]; [122].

Os estudos relacionados a pesquisas envolvendo a identificação e descoberta de novos fármacos, cujos constituintes ativos são plantas medicinais ou estratégias de aperfeiçoamento dos fitoterápicos, permanecem sendo de grande importância na atualidade. De acordo com os últimos estudos, no ano de 2020, mais de 50% dos fármacos decorrentes ou inspirados em produtos naturais foram aprovados pela agência americana FDA (Food and Drug Administration), assim como houve fármacos aprovados por instituições com a mesma finalidade [4]; [13].

No entanto, mesmo com a diversidade de plantas medicinais existente na flora, muitas espécies nunca foram identificadas e estudadas suas propriedades visando a descoberta de novos fármacos. É necessária uma ampliação na investigação das plantas desconhecidas e existente na flora, assim como as que já são utilizada tradicionalmente por diversas comunidades, pois essa exploração garante estudos que visionem o desempenho e seguridade no uso desses possíveis fármacos a serem cientificamente comprovado [12]; [13]; [14].

Os diversos produtos naturais fornecem propriedades peculiares em comparação com moléculas produzidas sinteticamente por pesquisadores de compostos sintéticas para finalidade terapêutica, o que acarreta em uma maior complexidade no processo de descoberta e desenvolvimento de novos fármacos [127]. A extensa diversidade de cadeia e complexidade estrutural são as principais características desses compostos. Normalmente, os produtos naturais possuem uma massa molecular mais elevada, uma maior quantidade de número atômico em seu carbono com hibridização sp^3 e oxigênio, além de tipicamente possuir uma menor número atômico de halogênio e nitrogênio, contudo apresenta elevados números de

doador- aceitador de ligações de hidrogênio, coeficiente de partição em sistemas octanol/água, calculados mais baixos indicando maior hidrofiliabilidade da molécula através do logaritmo (cLogP) e conseqüentemente proporcionando uma maior rigidez da molécula quando comparada à biblioteca de compostos químicos sintéticos. [128]; [129]; [11].

A rigidez molecular dos produtos naturais tende a possuir grande vantagens como alvos terapêuticos na descoberta de fármacos em que aborda interação do tipo proteína-proteína entendidas como o contato físico entre proteínas (duas ou mais), que tais interações ocorrem em um organismo vivo ou células [130]. Os produtos naturais possuem compostos promissores para descoberta de fármacos orais, além de adequar a “Regra dos Cinco” de Lipinski, em que estabelecem que 4 ou 5 fatores sejam apresentados nos compostos, concretizando uma das regras de boa biodisponibilidade oral teórica [131]; [132]. Nos últimos anos, muitos dos medicamentos desenvolvidos não atenderam a essa regra ilustrada pelo alto peso molecular das drogas orais aprovadas [133]. Os produtos naturais evoluíram ao longo do tempo, até se tornarem estruturas moleculares “otimizadas”, objetivando principalmente para servir a funções biológicas especificamente privadas [134], principalmente visando padronização de mecanismos dos agentes endógenos de defesa e por vezes a interação ou competição com diversos organismos, o que demonstra sua alta importância para a prevenção e tratamento diversas doenças. Ademais, fatores relacionados as informações sobre eficiência, eficácia e seguridade, são elucidados mediante estudos da científico associado ao conhecimento popular. Os produtos naturais em geral, possuem um pool amplamente beneficiado com compostos bioativos responsável por envolver toda a área do espaço químico do composto natural [135].

O planejamento e desenvolvimento de fármacos oriundos de produtos naturais, possuem vantagens exitosas, no entanto existem algumas desvantagens que desmotivam a indústria farmacêutica em estudos e investigações através de programas de descoberta de drogas baseada em fontes naturais [136]. A realização dos estudos envolvendo produtos naturais, são baseadas em triagens compostas por diversos extratos de plantas medicinais [137], e que por vezes podem sofrer de incompatibilidade com ensaios biológicos baseados em alvo [138]. O método de identificação de compostos bioativos não é uma tarefa fácil, tendo em vista sua complexidade e que as ferramentas de estratégias de desreplicação constitui um desafio e sua ampliação é necessária para evitar a identificação de compostos descobertos anteriormente. Em geral, diversas vezes é desafiante realizar o acesso ao material biológico necessário para efetuar o isolamento e caracterização de um produto natural bioativo [139]. Outro desafio relacionado ao estudo de produtos naturais não modificados que apresentam compostos bioativos relevantes, são a obtenção dos direitos de proteção intelectual, tendo em vista que os compostos naturais

sem modificações e totalmente original são dificilmente patenteados a depender do quadro jurídico do país de depósito, entretanto, compostos derivados simples recebem a proteção através de patentes quando realizado o pedido pelo titular [140].

Os produtos naturais são um potencial fonte de novos agentes terapêuticos para o tratamento da Doença de Chagas, dentre outras doenças parasitárias [15]; [123]. No caso das plantas, as pertencentes à família Euphorbiaceae são uma importante fonte alternativa para o tratamento dessa doença causada pelo protozoário *T. cruzi*. São realizados muitos estudos por diversos autores, com extratos de Euphorbiaceae, com metabólitos das mais diversas classes químicas isoladas desses extratos [16].

Nos últimos anos, o desenvolvimento de novas técnicas de bioensaio, métodos de biotecnologia, estudos fitoquímicos bio-guiados, triagem automatizada de alto rendimento e métodos analíticos de alto desempenho introduziram novos conceitos e possibilidades de design racional e descoberta de medicamentos. Nesse contexto, os produtos naturais têm desempenhado um papel importante e decisivo no desenvolvimento da química medicinal moderna. A descoberta de novos ligantes oriundos de produtos naturais para novos antiprotozoários é um importante passo da pesquisa científica, tendo em vista que esses produtos são responsáveis por pelo menos 50% dos novos fármacos descobertos desde 1981 [4].

3.2 Família Euphorbiaceae

A família Euphorbiaceae é uma das mais abundantes e diversificadas em número de fanerógamos das angiospermas, em que sua ocorrência é mais significativa em ambientes tropicais e subtropicais em diversas partes do mundo. Em relação à ocorrência, há registros de aproximadamente 300 gêneros e 7600 espécies, e se consolida cientificamente como sendo uma importante fonte de substâncias biológicas ativas para assim efetivar a descoberta e desenvolvimento de medicamentos [17]; [18]; [19]; [20]; [21]. Amplamente utilizada na medicina popular, essa família é uma importante fonte de investigações científicas, visando a busca de seus constituintes químicos e bioativos.

A maioria das plantas pertencentes à família Euphorbiaceae são ervas, arbustos ou árvores, principalmente nos trópicos e subtropicais em diversos países do mundo. Muitas dessas espécies possuem semelhança com os cactos. As regiões onde são encontradas em maior abundância são nos continentes americano e africano, especialmente nas regiões Indomalásia e América tropical, onde existe uma maior concentração de espécies. Entretanto, diversos gêneros

e espécies dessa família de plantas são encontrados na África do Sul, Estados Unidos, ou seja, em regiões não tropicais, porém em quantidade reduzida em comparação com regiões tropicais [141].

As principais características morfológicas da família Euphorbiaceae são suas folhas alternadas, dificilmente opostas, e apresentam apêndices. Elas possuem aspectos simples, porém quando compostas, apresentam forma palmada. As estruturas do tipo estípulas presentes nessa família de plantas são compostas por espinhos, pelos e glândulas. Morfológicamente, as flores actinomórficas (radialmente simétricas) são unissexuais na mesma planta, ou seja, masculinas e femininas (polígamas). A diversidade estrutural das flores é esperada para as espécies de Euphorbiaceae, tendo em vista a amplitude e grandeza dessa família, em que muitas delas o sistema reprodutivo é classificado como monoicas (flores unissexuais) ou dioicas (flores diclinas). Os estames são definidos como órgãos masculinos responsáveis pela produção de pólen e as flores femininas possuem ovários superiores (hipóginas) [142]; [143].

Economicamente é considerável a quantidade de espécies da família Euphorbiaceae presente no mercado. Dentre elas, destacam-se a *Manihot esculenta* (mandioca), amplamente cultivada em regiões de climas tropical e subtropical do mundo e que possuem suas raízes comestíveis; a *Ricinus communis* (mamona); *Hevea brasiliensis* (seringueira e árvore-da-borracha), que possuem espécies utilizadas em ornamentações, como a *Acalypha arvensis* Poepp. (rabo-de-macaco); *Acalypha accedens* Mull.Arg. (acalifa); *Jatropha gossypifolia* (pião-roxo). Muitas ervas dessa família foram consideradas invasoras na América do Norte, dentre elas a *Euphorbia esula*, que possuem spurge folhoso e *Triadica sebifera* conhecida como sebo chinês. Na área medicinal, diversas espécies têm demonstrado, através de estudos científicos, boa eficácia para inúmeras enfermidades [144]; [145].

Dentre os diversos gêneros da família Euphorbiaceae, recebe destaque o Gênero Cróton, que possui em seu catálogo aproximadamente 1200 espécies no Brasil. Amplamente empregue por diversas comunidades na medicina popular, esse gênero é promissor em inúmeros estudos já realizados por diversos pesquisadores, principalmente em relação a seus óleos essenciais e a diversidade de substâncias ativas. Por isso, sua grande importância econômica e científica. Somados a esses fatores, alguns estudos já comprovaram atividade biológica por diversos cientistas pesquisadores, apontando o Cróton (figura 1) como um produto natural eficaz e um candidato promissor para a descoberta de novos medicamentos naturais [22]; [23].



Figura 1. Espécie do gênero *Croton* (*Croton heliotropiifolius*)

Fonte: <http://diacaatingaflora.blogspot.com/2013/07/velame-croton-heliotropiifolius.html>.

Acesso em: 09 de maio de 2023.

Outro gênero da família Euphorbiaceae que recebe destaque é a *Euphorbia* (figura 2), da tribo Euphorbieae, pertencente à subtribo Euphorbiinae. Apresentam caracteristicamente, no formato de pseudanthium, palavra de origem grega que significa "flor falsa". Conceitualmente, consiste em uma inflorescência que se assemelha a uma flor, ou seja, são gêneros com plantas que possuem diversas flores verdadeiras, denominada de cyathium [146].

As principais características dessa formação de inflorescência são a formação de um involúcro assemelhado a uma taça, que possui brácteas fundidas e glândulas nectárias periféricas, envolvendo um anel de flores masculinas, onde cada uma possui um estame único. Na região central do cyathium, está presente uma flor feminina na base do involúcro, composta por um ovário em um pequeno caule com pistilo de estigmas no formato ramificado. Toda essa estrutura é semelhante a uma flor única. Em relação aos frutos característicos das espécies desse gênero de plantas, recebem destaque os do tipo esquizocarpo. Estes são frutos secos e deiscentes, com diversos carpelos, geralmente mais de três células. No período de maturidade, eles se abrem espontaneamente em partes separadas e logo em seguida ocorre uma explosão que acarreta na separação e no espalhamento das sementes [147].



Figura 2. Espécie do gênero *Euphorbia* (*Euphorbia hirta*)

Fonte: <https://autordapropriasaude.com.br/planta/erva-de-santa-luzia/>. Acesso em: 09 de maio de 2023.

Diversos estudos abordam as potencialidades dos metabolitos secundários dos gêneros de plantas medicinais pertencentes à família Euphorbiaceae, incluindo alcaloides, compostos fenólicos, terpenoides, fenilpropanoides, dentre outros [24]. Os compostos orgânicos voláteis – COVs (óleos essenciais), provenientes de espécies dos gêneros dessa família de plantas, possuem grande importância por apresentarem inúmeras atividades biológicas, dentre elas a anti-inflamatória [24]; [25]; [26]; [27], antibacteriana, antifúngica [28]; [29]; [30]; [31], antioxidante [32]; [33]; [34]; [35]; [36], antiparasitária, antiulcerogênica [28]; [37].

Na literatura, são relatadas inúmeras investigações relacionadas a espécies da família Euphorbiaceae, dentre elas as estudadas por Lai e colaboradores, que utilizam plantas medicinais para a fitoterapia chinesa, tendo catalogado 33 espécies de 17 variados gêneros de Euphorbiaceae [38]. Dentre essa diversidade de espécies, várias são utilizadas no tratamento de picadas de cobras, realizando assim o combate 'toxina com toxina'. Outras são utilizadas como plantas ornamentais, como a *Euphorbia tirucalli*, e outras de grande relevância econômica, como a *Manihot esculenta*, *Ricinus communis* L., *Jatropha curcas* L. [19]

A diversidade de espécies da família Euphorbiaceae é conhecida por apresentar em sua composição fitotoxinas (substâncias tóxicas presentes em diversas plantas), sendo os gêneros *Jatropha* e *Euphorbia* apontados na literatura como venenosos, podendo inclusive induzir a distúrbios na pele [20]. Dentre as espécies do gênero *Jatropha*, a *J. curcas* L. é uma das que apresenta grande toxicidade, principalmente para microrganismos humanos, além de alguns animais [39]. Algumas espécies apresentam propriedades econômicas, como o látex, possibilitando a produção de borracha natural. O látex contém alguns compostos, incluindo

triterpenoides, diterpenoides, enzimas, entre outros. Na família Euphorbiaceae, o látex é detectado especialmente nas subfamílias Crotonoideae e Euphorbioideae [40].

Dentre a diversidade de plantas medicinais da caatinga, as pertencentes à família Euphorbiaceae são amplamente utilizadas por diversas comunidades tradicionais para uso terapêutico. Essas plantas são produtoras de um diversificado número de metabólitos secundários. As plantas dos gêneros *Croton* e *Astraea* recebem destaque nessa família, pois essas representantes possuem importância econômica rica em constituintes ativos, representados por metabólitos secundários empregados em diversos estudos, como biossíntese, quimiotaxonomia, etnobotânica, ensaios biológicos e/ou farmacológicos, modelos para síntese orgânica, entre outros [16].

Estudos fitoquímicos realizados com espécies da família Euphorbiaceae descrevem a ocorrência de compostos com grande diversidade estrutural, tais como a acetofenona xantoxilina, esteroides, triterpenos, cumarinas, flavonoides, derivados fenólicos, diterpenoides e alcaloides, entre outros. Com efeito, os terpenos constituem o maior grupo de metabólitos secundários derivados das plantas (com mais de 40.000 moléculas conhecidas). Dentre esses, destaca-se o ascaridol, uma substância orgânica de origem natural pertencente à classe dos monoterpenos bicíclicos, que possui uma ponte formada pelo grupo funcional peróxido, encontrada principalmente em óleos essenciais [41]; [124]. Em resumo, de acordo com alguns estudos, o monoterpeno ascaridol apresenta-se como um promissor agente fitofármaco com grande potencial de atuação em propriedades biológicas antiprotozoárias. São necessários mais estudos que abordem o ascaridol como substância ativa com múltiplos potenciais farmacológicos [42]; [125]; [126].

3.3 Doença de Chagas

Históricamente, o *Trypanosoma cruzi* (figura 3) está presente no mundo desde a antiguidade, período em que os continentes estavam juntos em um único bloco. O parasita fazia-se presente no sistema digestivo de animais aquáticos, répteis, entre outros. Ao longo dos anos, o parasita passou por diversas evoluções, chegando a desenvolver seu ciclo em alguns insetos e em pequenos mamíferos. Inicialmente, esse parasita iniciou seu ciclo de vida em insetos vetores (pernilongos, pulgas, mosquitos, entre outros), e com a evolução expandiu-se ao homem. Tempos mais tarde, a tripanossomíase disseminou-se integralmente na América e, na sequência, estabeleceu-se através do ciclo humano como uma infecção parasitária que aflige diversas sociedades até os dias atuais [43]; [44]; [45].

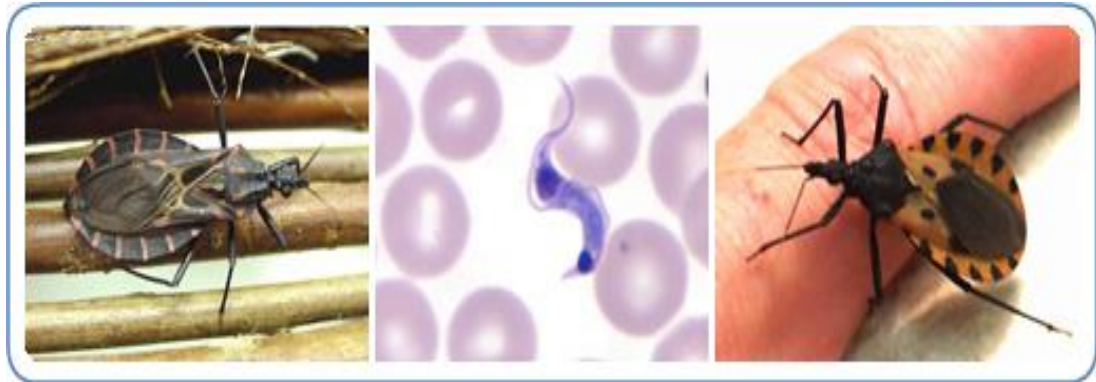


Figura 3. Parasita *Trypanosoma cruzi*

Fonte: <http://www.cdc.gov/parasites/chagas/> Acesso em: 09 de maio de 2023.

A Doença de Chagas (DC) é uma infecção causada pelo parasita *Trypanosoma cruzi*. A DC é uma das doenças negligenciadas mais relevantes endêmicas em diversos países da América Latina. O modo habitual de transmissão dessa doença dá-se por meio de vetores, consistindo em aproximadamente 80% dos eventos de contágio da patologia, e ocorre através do contato do homem com os excrementos contaminados do inseto (do gênero *Triatoma*), que é o vetor. A infecção por essa doença em humanos e outros mamíferos ocorre principalmente pela via vetorial, pois ao nutrir-se por intermédio de um dos diversos hospedeiros existentes no meio ambiente, contagiados com as fezes de insetos hematófagos, ocorre a propagação da doença [1]; [50]; [51].

A transmissão do *T. cruzi* através da mãe para o filho possui uma estimativa de casos de aproximadamente 4,7%, com intervalos variando de 3,9 a 5,6%. Essa taxa pode ser ainda maior em países considerados endêmicos, em comparação com regiões não endêmicas, onde o índice de transmissão estimado é de aproximadamente 5,0 a 7,0% [148]. A transmissão vertical congênita possui como principal determinante biológico a parasitemia ou viremia materna, com taxas de aproximadamente 31%. Em situações em que o *T. cruzi*, o agente causal da doença de Chagas, é detectado, a quantificação da carga parasitária é realizada através do exame de PCR (Reação em Cadeia da Polimerase) em pacientes chagásicos crônicos. A transmissão pode ocorrer mesmo quando o resultado do PCR é negativo [149].

Um estudo comparativo foi sugerido entre as DTUs (Unidades Discretas de Tipagem), taxas reduzidas de infecção e transmissão de crianças que nasceram de mães apresentando sorologia positiva para *T. cruzi*, como a TcII, sendo comparada com a TcV no Brasil. Em contrapartida, é estimado que a resposta imune celular na infecção pelo vetor desempenha um importante papel nas infecções perinatais. A exposição sustentada à infecção pelo *T. cruzi* apresenta diversas reações que ocasionam a identificação do organismo parasitário,

proporcionando uma resposta imune devido à frequência da exposição ao vetor infectado. Isso pode permitir o monitoramento do desenvolvimento do parasita durante todo o ciclo de vida do indivíduo [150].

A infecção da doença de Chagas através do sangue e por meio de medicamentos produzidos pelo fracionamento industrial do plasma humano (Hemoderivados) possui uma estimativa de transmissibilidade taxada em aproximadamente 10 a 25% por unidade de sangue infectado. O número relacionado às taxas de infecções através do transplante de órgãos com doadores infectados possui redução nos índices em receptores renais, onde as taxas variam de 0 a 19%, em comparação com receptores hepáticos (fígado) com taxas de 0 a 29%, e em receptores de coração, as taxas variam de 75 a 100%. A forma de transmissão menos frequente é a via oral, através do manuseio e consumo de alimentos ou bebidas contaminadas por triatomíneos ou até mesmo suas fezes [153]; [154].

O ciclo de vida do *T. cruzi* em seres humanos ocorre de duas formas evolutivas morfológicas. A tripomastigota, que corresponde à forma sanguínea da infecção, caracteristicamente apresenta um flagelo alongado livre que se estende internamente percorrendo o corpo do parasita, formando uma membrana ondulante. Isso não ocorre através de sua divisão diretamente no sangue; no entanto, facilita sua movimentação pela corrente sanguínea, disseminando a infecção internamente por todo o corpo do infectado. O formato amastigota, de maneira geral, é formado em células musculares. Eles possuem formato arredondado e caracteristicamente não possuem flagelo. Sua multiplicação e disseminação corporal ocorrem em diversas células, recorrentemente optando pelas células-tronco mesenquimais (figura 4) [155]; [156]; [157].

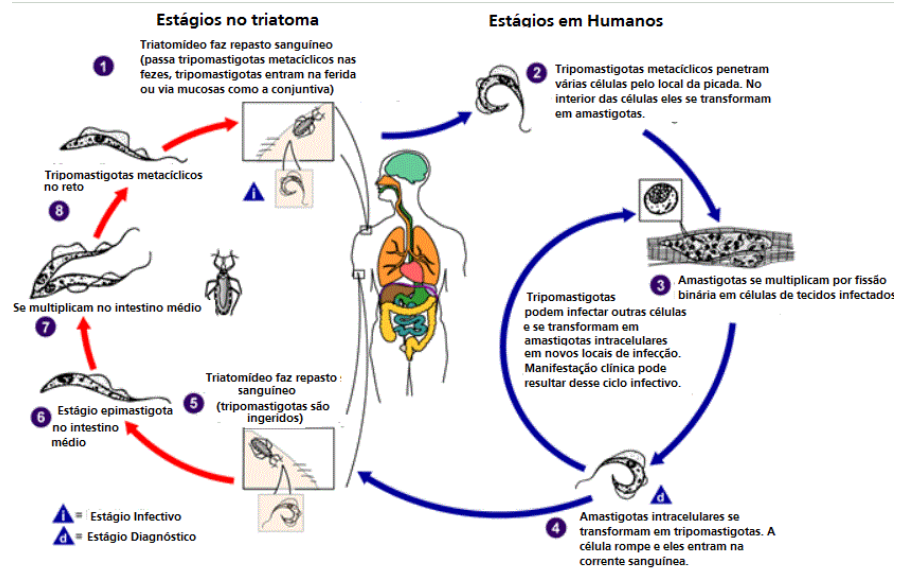


Figura 4. Ciclo de vida do *T. cruzi*

Fonte: <https://www.cdc.gov/dpdx/>. Acesso em: 15 de junho de 2023.

As populações de *T. cruzi* possuem grande heterogeneidade genética e diversidade fenotípica dentro do táxon, com uma imensa importância biológica. O parasito está presente em diversas espécies de vetores e hospedeiros, circulando em inúmeras variedades de mamíferos, desde os silvestres até os domésticos, entre outros, de diversas características fitogeográficas. Recebem a classificação de seis DTUs (Unidades Discretas de Tipagem), que são nomeadas TcI (TcI a TcVI) [158]; [159].

A ampla diversidade da variabilidade genética do parasito tem sido associada a fatores geográficos em termos de distribuição, assim como à resposta terapêutica, patogenia e ao perfil clínico do parasitado [160]. A transmissão do *T. cruzi* ocorre em áreas endêmicas deste protozoário, por meio de espécies pertencentes a três gêneros de triatomíneos, popularmente conhecidos como bicho-barbeiros ou simplesmente "barbeiros" (*Panstrongylus* sp., *Triatoma* sp., *Rhodnius* sp.). Esses gêneros estão distribuídos de maneira diversa em países da América Latina, da Argentina ao Chile, entre outros, além de regiões formadas por florestas e com clima seco [161]; [162].

A doença de Chagas apresenta em seu curso clínico bifásico uma fase aguda, que em muitos casos é assintomática, podendo evoluir para a fase crônica. Os casos de infecção aguda acometem principalmente crianças nos primeiros anos de vida. No entanto, podem ocorrer em qualquer idade. Em diversos casos, manifestam-se de forma assintomática. Entretanto, quando apresentam sinais e sintomas, os principais são febre prolongada, inoculação subcutânea com inflamação no local, formação de edema periocular unilateral nas pálpebras (sinal de porta de

entrada "chagomas de inoculação"), adenopatia, hepatomegalia e esplenomegalia. A duração da fase aguda é de cerca de 4 a 8 semanas e, aproximadamente, 90 dias após ocorre a redução da parasitemia [163]; [164]. Na fase aguda grave, ocorre em uma pequena proporção de pacientes, entre 1 e 5%. Os principais sinais e sintomas são miocardite, que são distúrbios do coração e vasos sanguíneos, além de pericardites, entre outros [43].

As manifestações clínicas em recém-nascidos, que ocorrem devido à infecção congênita, na maioria das vezes são assintomáticas ou com sintomas moderados. Em uma pequena parcela dos casos nesse público, ocorre gravidade extrema com risco de óbito [166]. Surto de infecção causada por transmissão oral por alimentos ou bebidas contaminadas com fezes de triatomíneos apresentam maior índice de mortalidade em comparação com a transmissão vetorial tradicional [167].

A infecção aguda da DC normalmente pode ter sua resolutividade espontânea, porém, após a fase aguda, os indivíduos infectados frequentemente evoluem para a fase crônica, caso não recebam tratamento prévio. O maior grupo de pacientes afetados por infecções do parasito não desenvolve sintomas ou apresenta ligação visceral. A forma indeterminada da doença de Chagas possui um prognóstico considerado bom e soropositividade para infecção por *T. cruzi* nas manifestações clínicas de pacientes diagnosticados com a doença, porém, na ausência de sinais e sintomas clínicos que causem comprometimento radiológico e acometimento que ocasionem complicações no funcionamento cardíaco ou digestivo [168]. Indivíduos acometidos com as formas indeterminadas dessa doença estão sendo recorrentemente detectados com anomalias cardíacas, possivelmente devido a métodos de diagnóstico mais precisos como a ressonância magnética (RM) e a ecocardiografia, entre outros [169].

A infecção chagásica causa danos secundários aos órgãos e apresenta resposta inflamatória na fase aguda, relacionados à ação do hemoflagelado *T. cruzi* diretamente. A forma amastigota do parasito pode ser encontrada comumente em diversos tecidos, incluindo o esquelético e o músculo liso, além de outros locais como o sistema fagocítico mononuclear (SFM), gônadas, entre outros [170]. Na fase amastigota, os mecanismos de resposta a infecções e a avaliação do controle com ampla eficiência contra o *T. cruzi* correspondem às etapas de uma resposta inflamatória intensificada, através da produção de anticorpos após contato com o parasito, associada à resposta imunológica natural ativada por células NK (Natural Killer) e macrófagos, através do fator de necrose tumoral-alfa (TNF- α) e interferon-gama (IFN- γ), que correspondem às citocinas Th1 [171].

O diagnóstico parasitológico do contágio por *T. cruzi* é realizado mediante análise de microscopia óptica de tripomastigotas sanguíneos ou através de exames utilizando líquido

cefalorraquidiano – LCR (fluidos corporais) [172]. Em casos de infecções em que a doença é transmitida da mãe para o bebê durante o período de gestação (congenita), o exame para o diagnóstico pode ser baseado em IgG anti- *T. cruzi*. A infecção chagásica congênita pode ser diagnosticada através de um exame de sangue tradicional ou por esfregaços de sangue, que são exames laboratoriais eficazes e de baixo custo [173]. Existem outros métodos de identificação dos parasitas, como o teste de Strout e a micro-hematócrito centrifugação do sangue, que possuem maior sensibilidade e rendimento. Os métodos indiretos para diagnóstico parasitológico detectam a presença do parasita, tais como o PCR (reação em cadeia da polimerase), mais sensível que o xenodiagnóstico e a hemocultura, que possuem menor sensibilidade [174].

A patologia da DC possui agentes etiológicos parasitas taxonomicamente relacionados com particularidades morfológicas e bioquímicas análogas. Os fármacos geralmente empregados no tratamento dessa patologia apresentam problemas de toxicidade, efetividade inconstante, formalidades para gestão e/ou observância de regulamentos de terapia parenteral. A doença de Chagas apresenta poucos medicamentos ou tratamentos em desenvolvimento clínico [1].

A doença de Chagas é endêmica em países do continente americano, em territórios de clima tropical, afetando as populações em situação de pobreza ou pobreza extrema, desigualdade em saúde. Por outro lado, atualmente ocorrem casos em áreas não endêmicas, com elevado desenvolvimento socioeconômico [93]; [94]. A ocorrência da doença acarreta deformidades e incapacidades físicas, podendo causar morbimortalidade. Outro fator agravante decorrente dessa doença é o sofrimento, estigma e preconceito que envolvem a doença. Assim, proporcionar o desenvolvimento humano é uma ferramenta essencial para alcançar o sucesso social [93]; [95].

No Brasil, a região nordeste se destaca como a segunda principal região endêmica da doença de Chagas. A Bahia é um dos estados que possui uma maior diversificação de espécies de triatomíneos no país. O estado do Ceará também se destaca em relação ao elevado número de casos, principalmente na região sudoeste do estado, com cobertura vegetal predominante de caatinga. Isso ocorre porque possui uma vasta área rural e habitações humanas precárias encontradas na zona rural do estado, facilitando a presença de diversas espécies de triatomíneos e a transmissão da doença de Chagas [96]; [97].

A DC é uma enfermidade patológica classificada como doença tropical negligenciada, que tem recebido pouco atenção por parte do governo e da indústria farmacêutica. Nos últimos

anos, o investimento global em medicamentos antiparasitários tem sido apenas 0,1% do capital universal empregado [46]; [47]; [48].

O combate a tripanossomíase é considerado desafiadora para muitos pesquisadores, principalmente devido a seu sistema de transmissão possuir uma grande complexidade que envolve transmissão congênita, oral, transfusão de sangue, além de envolver os vetores [52]. No Brasil a principal forma de transmissão é via oral, tendo em vista a precariedade de muitos bairros periféricos sem saneamento básico e muitas residências sem as mínimas condições básicas que servem de habitat para o barbeiro. Outro modo de transmissão é ingestão de alimentos contaminados pelo vetor transmissor da DC. O principal mecanismo de contenção para a DC é a vigilância sanitária, por ser uma doença infecciosa transmitida entre animais e pessoas (zoonose), pois essa enfermidade ainda não possui um tratamento eficiente [53].

Atualmente, apenas dois medicamentos são utilizados para tratar a doença de Chagas: o Nifurtimox e o Benznidazol (figura 5). No entanto, eles possuem eficácia reduzida na forma crônica. Dessa forma, é perceptível a necessidade urgente de novos medicamentos capazes de tratar as formas evolutivas de *T. cruzi* [54]; [55]. Os medicamentos antitripanossômicos são comumente recomendados para a doença de Chagas na fase aguda e congênita. Entretanto, existem estudos que divergem em relação à eficácia do tratamento etiológico para essa doença realizado na fase aguda. Os resultados dessa metanálise apontam que o fármaco antichagásico Benznidazol (BNZ) possui poucos benefícios com base no estudo BENEFIT, pois o seu uso não resultou na redução do estado clínico cardíaco de indivíduos que possuem cardiomiopatia em estados moderados ou com maior gravidade. No entanto, em mulheres que passaram por tratamento antichagásico, foi observada uma ampla eficácia terapêutica, promovendo a interrupção da transmissão da mãe para o filho durante a gestação ou parto [175]; [176].

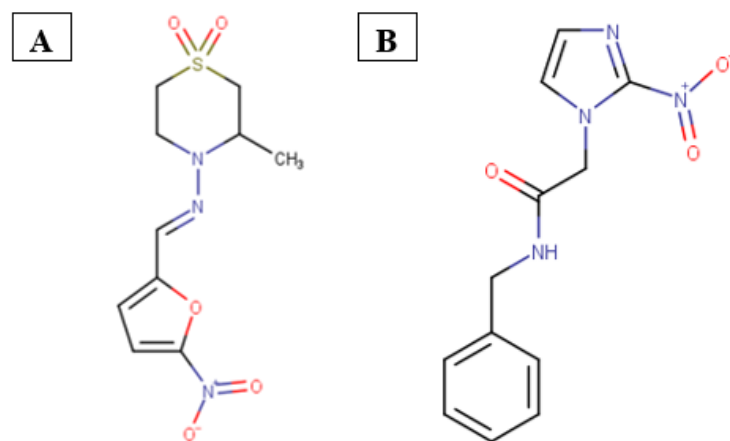


Figura 5. Estrutura 2D do Nifurtimox (A) e Benznidazol (B)
Fonte: GARCÍA-HUERTAS et al., 2021^[54]

Dentre os grandes desafios na pesquisa de agentes quimioterápicos para a doença de Chagas, assim como de outros novos fármacos para outras doenças negligenciadas, podem-se destacar a motivação aos químicos medicinais a pesquisar na área, assim como a validação de novos alvos moleculares e a busca de novos compostos líderes. E eminentemente o incentivo da indústria farmacêutica, especialmente a brasileira, estimulando a importância do desenvolvimento dessa classe de fármacos. Nesse sentido, o tratamento contra estas doenças ainda constitui um papel desafiante para a ciência, estimulando diversos grupos de pesquisa à descoberta de novos medicamentos mais efetivos e menos tóxicos [98].

Sendo a doença de Chagas vista como uma das enfermidades mais negligenciadas, diversos estudos são realizados com o intuito de indicar caminhos para a renovação no tratamento das mesmas. Produtos naturais são vistos como uma ampla possibilidade, tendo em vista que diversas plantas têm revelado efeito terapêutico contra a doença de Chagas. Os produtos à base de plantas são possíveis “minas” imensamente diversas de compostos com propriedades terapêuticas [49].

3.4 Alcaloides

Os alcaloides derivam de plantas, animais, bactérias, entre outros organismos, que produzem esse metabólito secundário. Esses compostos possuem ampla relevância para a terapêutica, principalmente os de origem vegetal, por serem substâncias caracteristicamente eficientes e estruturalmente diversificadas [177]. Existem aproximadamente 5.500 alcaloides conhecidos, além de muitos outros que continuam sendo isolados e descobertos por pesquisadores e grupos de pesquisa em todo o mundo, com o objetivo principalmente de obter novos fármacos [56]. O significado da palavra alcaloide é "semelhante aos álcalis", derivado de "alcalino", que caracterizava qualquer base contendo o elemento químico nitrogênio. Os grupos vegetais angiospermas, ou plantas com flores, são as principais fontes naturais de alcaloides; esse metabólito promove nas plantas a defesa contra herbívoros (insetos) e patógenos (doenças vegetais). Diversos alcaloides são usados na terapêutica desde a antiguidade, por meio de extratos vegetais no tratamento de enfermidades, como picadas de cobra, febres intensas, e nos dias atuais, permanecem sendo substâncias altamente relevantes e promissoras para a medicina [179]; [180]."

Os metabólitos secundários alcaloides são compostos provenientes das aminas. De maneira geral, esses metabólitos são aminas que possuem em sua estrutura química uma cadeia fechada com a presença de nitrogênio entre carbonos do ciclo presente nas plantas. Essa

composição é responsável pelo sabor amargo das plantas, e muitas delas causam dependência física e psíquica [56]. A principal característica desse composto é a existência de hidrogênio, átomo que possibilita que o composto possua propriedades farmacológicas [57]. A principal fonte de alcaloides são as plantas medicinais, nas quais esses metabolitos são comumente encontrados nos caules, cascas e raízes das plantas, e diversos compostos já foram identificados [58]; [59].

Estudos químicos demonstram que as plantas que contêm alcaloides são ricas em princípios farmacológicos, principalmente antidepressivos, antibacterianos, neuroprotetores (no tratamento da doença de Alzheimer), dentre outros [60]; [61]; [62]. As investigações envolvendo alcaloides como fármacos são amplamente relevantes, principalmente devido à sua alta toxicidade para humanos e animais. Diversos estudos comprovam que a estrutura química, diversidade e mecanismos de ação são determinantes para ocasionar efeitos tóxicos de diversos alcaloides [63]; [64]. Entretanto, os efeitos de toxicidade desses metabolitos dependem de diversos fatores, dentre eles o tempo de exposição, o local de ação, a dosagem, a sensibilidade, dentre outros [65].

Os alcaloides presentes nas plantas medicinais são compostos que possuem nitrogênio envolvido para proteção contra pragas predadoras, além de atuar como reguladores do crescimento e suprir alguns minerais essenciais [66]. Quando os predadores agridem severamente as plantas, isso acarreta na interferência do sistema nervoso das pragas, na síntese de proteínas e no transporte de membrana [67]. Em relação ao estresse biótico continuamente exposto, causado por organismos vivos, e ao estresse abiótico, causado por irradiação e temperatura, há um aumento no índice reprodutivo das plantas fitoterápicas [68].

A utilização histórica de alcaloides como fármacos demonstrou a relevância desse metabolito secundário para a indústria farmacêutica. Diversos alcaloides são utilizados como antibacterianos e antiparasitários, como os oriundos do grupo β -carbolinas, como a harmina [69]. Vários alcaloides são testados contra o *T. cruzi* e esses estudos demonstram notável atividade anti-*T. cruzi*, principalmente nas formas epimastigotas, que foram relacionadas aos inibidores que paralisam a respiração pela interrupção da cadeia respiratória (inibição da respiração celular). A apomorfina é um dos compostos já estudados que apresenta um grande potencial em atividade antichagásica em testes *in vivo* [70]

Os alcaloides possuem diferentes classificações, dentre eles as botânicas, que se dividem em alcaloides da "quina" e do "ópio", além da classificação biossintética, dividida em alcaloides verdadeiros oriundos de um aminoácido e que possuem anel heterocíclico com átomo de

hidrogênio. Os protoalcalóides são compostos químicos que possuem um átomo de nitrogênio oriundo de um aminoácido e que não pertence ao heterocíclico do composto orgânico [181].

A classificação dos alcaloides apresenta dificuldades, tendo em vista que as moléculas orgânicas das estruturas químicas são amplamente complexas [71]. Diversas investigações envolvendo alcaloides determinam as distintas maneiras de classificar esses metabólitos secundários, como origem bioquímica, efeito farmacológico, origem taxonômica e estrutura química. A origem bioquímica e a estrutura química são os sistemas de classificação mais utilizados regularmente [56].

Os alcaloides são sintetizados naturalmente pelas plantas, com base em sua precisão. Alguns desses metabólitos secundários são documentados como causadores de toxicidade em humanos e animais, como indolizidina, hepatotóxicos e tropano [72]; [73]. Em humanos, esses alcaloides ocasionam intoxicação através da ingestão e muitas vezes pela inalação. Além disso, outro método capaz de causar efeitos tóxicos é por meio da utilização de produtos naturais que possuem alcaloides tóxicos em sua composição. Para os animais, a intoxicação ocorre em pastos e em locais com escassez de alimentos, facilitando assim a ingestão de plantas tóxicas [69].

Os metabólitos secundários: alcaloides tóxicos presentes nas plantas, causam sérios problemas quando consumidos pelo ser humano, podendo causar diversos malefícios ao corpo humano, mesmo quando a ingestão do extrato da planta seja em quantidade moderada. O nível da resposta à intoxicação do animal-alvo depende da sensibilidade do animal [63]. Os alcaloides tóxicos que constituem diversas plantas medicinais, ao serem consumidos por humanos, podem comprometer diversas funções corporais por meio de mecanismos biológicos que envolvem material genético em células e tecidos específicos, além de envolver receptores e enzimas [69]. Os principais sintomas ocasionados pela ingestão de alcaloides tóxicos são náusea, vômito, coceira, podendo avançar para efeitos crônicos como paralisia, arritmias, dentre outros, podendo levar à morte [72]; [63].

3.5 Triagem Virtual

A abordagem da triagem virtual foi inicialmente proposta na década de 90, precisamente no ano de 1990. Essa ferramenta possibilita descrever a utilização de modelos computacionais e algoritmos, com a finalidade de identificar novas substâncias moleculares bioativas [74]. Esse método é comumente definido como um processo de análise conjunta de diversas moléculas presentes em um banco por intermédio de diversas técnicas computacionais. Através desse processo, é possível identificar novas moléculas potencialmente ativas, assim como remover

moléculas com propriedades desvantajosas no quesito de poder causar efeitos bioquímicos, fisiológicos e moleculares no corpo humano, ou moléculas tóxicas [75].

A triagem virtual se popularizou bastante dentro do processo de ampliação crescente do âmbito de publicações e citações, sendo amplamente usufruída pela indústria farmacêutica. Isso se deve ao aumento do número de bancos de dados de estruturas de compostos e ao volume de dados contido nessa ferramenta, que se tornou fundamental. A triagem virtual é uma alternativa rápida e de baixo custo para a triagem e seleção de *hits*, permitindo ranquear ou filtrar uma série de compostos de bancos de dados utilizando um ou mais procedimentos computacionais. Isso pode ser utilizado para indicar quais compostos devem ser sintetizados ou comparados, reduzindo o número de compostos selecionados para avaliação experimental [76].

A triagem virtual possibilita uma diversidade de aplicações e técnicas, visando principalmente a otimização, além da identificação de compostos inéditos e promissores até os dias atuais [77]. A aplicabilidade dessa ferramenta viabiliza o início da descoberta e desenvolvimento de novas drogas, podendo ser trabalhada em conjunto com a química combinatória e, especialmente, a Triagem de Alto Desempenho (High-throughput screening - HTS), que faz o uso de grandes coleções de compostos, principalmente na indústria farmacêutica, para a descoberta de medicamentos. Além disso, essa ferramenta é utilizada por diversas instituições e grupos de pesquisa, além de empresas de biotecnologia, nos casos em que o descobrimento de moléculas não é coberto pela técnica de triagem de alto desempenho HTS [78].

As técnicas de triagem virtual são baseadas tanto na estrutura como no ligante, indispensáveis para a descoberta de medicamentos inéditos, através da investigação de diversas bibliotecas moleculares. Para as baseadas em alvos (Target-Based Virtual Screening – TBVS), é indispensável um conhecimento amplo sobre a estrutura tridimensional dos alvos para estimar as probabilidades de interação, enquanto para a baseada em ligantes (Ligand Based Virtual Screening – LBVS), os descritores dos próprios ligantes são utilizados como propriedades discriminativas nos algoritmos de classificação e seleção, onde é feito o uso de um ligante conhecido ao menos, assim como sua ação biológica. Um método eficiente na utilização dessa técnica é o uso de bancos de compostos e suas atividades biológicas, além de criar sistemas matemáticos por intermédio dos descritores moleculares, oriundos dos ligantes, que posteriormente serão aplicados na triagem virtual [79].

Outra possibilidade estratégica envolvendo essa técnica está relacionada à triagem em bases de dados de compostos, que consiste em estudar através de adequação em sítios farmacofóricos de grande relevância para a triagem por similaridade e para ação biológica. Isso

ocorre visto que diversos compostos com forma estrutural semelhante possuem uma ampla possibilidade de apresentar características similares em suas propriedades [79].

De maneira geral, a triagem virtual corresponde à caracterização e classificação de compostos baseada em suas características, priorizando as que explicam com melhor clareza a ação biológica a ser estudada. Essas características norteiam compostos análogos que possuem possíveis atividades biológicas. Esse modelo de pesquisa é chamado de planejamento racional de fármacos, comumente utilizado para estudos na área de química medicinal, buscando contribuir na redução do tempo e custo de estudos em bancadas laboratoriais para sintetizar novas substâncias bioativas [82].

A predição de riscos de toxicidade, método este incluído na triagem virtual, é empregada na análise das características de compostos individualmente analisados, priorizando seus fragmentos estruturais e seus metabolismos de cada composto bioativo. Isso objetiva estudos com maior integralidade para a descoberta de possíveis medicamentos. Esse método é bastante relevante e indispensável em pesquisas na área da química de fármacos [83]; [84].

O método computacional de triagem virtual é uma ferramenta de grande importância no planejamento de novos fármacos, tendo em vista que através desse método é possível executar a seleção de compostos orgânicos por meio de bancos de estruturas moleculares com possíveis atividades biológicas, podendo ser inibidores enzimáticos ou antagonistas e agonistas de receptores, através do método *in silico* [83].

3.6 Predição de Dados

A predição de dados é fundamentada na classificação de modelos de dados. Esse modelo preditivo visa propor a previsão e possíveis tendências de interferência nos dados através de simulações em programas quimioinformáticos. Diversas técnicas são empregadas na predição de dados, dentre elas a curva ROC (Receiver Operating Characteristics), com a finalidade de selecionar, organizar e visualizar classificadores. A curva ROC permite a análise de gráficos em que são apresentadas as taxas de acerto e erros, além da verificação do desempenho de modelos classificatórios, possibilitando a comparação da eficiência dos sistemas classificatórios empregados [88].

Modelos classificatórios preditivos permeiam diversas probabilidades de uma instância ser de uma determinada classe, em que várias limiares poderão ter aplicação na geração de inúmeros conjuntos de saída, dessa forma podendo aplicar um limiar para o deslocamento da binarização. Esses estimadores estatísticos são aplicados na análise de desempenho de modelos

classificatórios, levando em consideração algumas amostras em que um modelo de dados recebe classificação segundo sua taxa de sensibilidade. Nesse caso, se uma instância é classificada como positiva, então essa é a fração dos verdadeiros positivos; caso seja classificada como negativa, ou seja, a fração dos falsos positivos (1-especificidade), baseado nos vários valores do teste realizado [88]; [89].

Diversas equações apresentam as métricas que são utilizadas para calcular com base na matriz de confusão. Dentre elas, a acurácia é conceituada por apresentar a proporção de predições corretas, considerando o acerto total e não apenas o que é positivo ou negativo. Outras equações são definidas como sensibilidade, especificidade, verdadeiro positivo preditivo e verdadeiro negativo preditivo [90].

O Coeficiente de Correlação de Matthews (MCC) é um método possível para realizar a avaliação global do modelo com base nos resultados alcançados na matriz de confusão. Esse método caracteriza-se como um coeficiente correlacionado aplicado em classificadores e modelos de predição binária, e como uma correção pode haver variações de $-1 \geq MCC \leq +1$. Quando o valor for 0 (zero), é classificado como aleatório. Dessa forma, para o valor ser classificado como confiável, é preciso que apresente valores próximos de 1,0 (um) [90]; [91].

Na curva ROC, além de ser possível analisar os pontos da curva, é viável a utilização de um indicador de dimensão dos efeitos. Um método frequentemente empregado nessas análises é a Área Sob a Curva (AUC), ou seja, a área abaixo da curva ROC. A AUC é definida como o resultado da integralidade dos pontos presentes no trajeto da curva ROC, que simultaneamente computa a especificidade além da sensibilidade, servindo como um estimador comportamental da acurácia do teste global [92]. Os valores referentes à AUC variam de 0 a 1, sendo classificados como uma porção dos valores unitários. Estatisticamente, uma AUC equivalente a 0,7 representa uma classificação de 70%. A interpretação dos valores de AUC é a seguinte: 0,5-0,6 (péssimo), 0,6-0,7 (ruim), 0,7-0,8 (pobre), 0,8-0,9 (bom), > 0,9 (excelente). Em geral, a curva ROC objetiva investigar modelos preditivos com uma elevada taxa de sensibilidade e um pequeno índice de falsos positivos [89].

3.7 Docagem Molecular

O docking molecular, ancoragem molecular ou docagem molecular é uma ferramenta essencial para predizer o melhor ajuste e orientação de um ligante a uma proteína, em que, quando o acoplamento é efetivado, forma-se um complexo estável. Em geral, é empregado em proteínas que possuem receptores e enzimas. Com base na orientação molecular, é possível

prever a afinidade e a força da associação de ligação entre o ligante e o receptor, ou seja, das duas moléculas em estudo [99].

A popularização da docagem molecular se efetivou inicialmente para o screening de compostos moleculares com possível potencial de atividade que interessava ao mercado industrial farmacêutico. No princípio, a principal finalidade era investigar a ligação entre compostos moleculares de pequeno e grande porte. No entanto, ao passar dos anos, transformou-se largamente, possibilitando sua utilização visando auxiliar diversas tarefas de programas que objetivam o descobrimento de potenciais medicamentos [100].

Os métodos computacionais de modelagem molecular são amplamente empregados em estudos de quimioinformática, objetivando principalmente representar compostos estruturais de forma mais próxima à realidade por meio de cálculos. Utilizam como instrumento matemático as bases da química teórica, visando a análise da capacidade antioxidante de compostos estruturais orgânicos e levando a compreender o mecanismo de ação antioxidante, incluindo os aspectos termodinâmicos e cinéticos. Dessa forma, a computação gráfica possibilita o manuseio de modelos obtidos, assim como a facilitação da interpretação e compreensão mais eficiente da relação biológica de estrutura-atividade de compostos [101].

Estudos relacionados ao campo da química teórica são extremamente relevantes para diversos estudos, pois essas ferramentas consistem na aplicação da técnica de modelagem molecular, possibilitando o diagnóstico conformacional de complexos moleculares [102]. A análise da interação de moléculas orgânicas entre proteína e o fármaco é uma investigação de alta relevância realizada nessa área de estudo. Os métodos clássicos e quânticos são as classes utilizadas para aplicação dessa teoria [103].

O desenvolvimento e aplicação de métodos químicos quânticos são essenciais para adquirir energia por meio da resolução da equação matemática do método de Schrödinger. Os métodos clássicos são desenvolvidos por meio das interações realizadas entre os componentes presentes no sistema, usando as leis da física clássica que possibilitam, por meio dessa equação, a obtenção do sistema no espaço de fase, definindo em detalhes toda a sua trajetória. Isso permite calcular as diversas propriedades termodinâmicas presentes por meio da mecânica estatística, sendo essa a principal ferramenta para a execução dos estudos [103].

As moléculas bioativas são compostas por grupos funcionais que envolvem a energia de ligação gerada entre o fármaco (ou micromolécula) e o sítio ativo, além da interação do fármaco com o receptor (alvo). Isso inclui a análise da conformação tridimensional regular assumida pelos segmentos esqueléticos moleculares proteicos [104].

A identificação tridimensional estrutural de compostos proteicos (proteínas) pode ser analisada por meio da técnica de raio X na cristalografia, além de possibilitar a identificação da posição atômica ocupada no polímero das proteínas, essencial para investigar sua função biológica. A técnica de cristalografia possibilita a realização do desenho de diversos fármacos por meio do método de ancoragem molecular (A.M.). Essa técnica possibilita o desenho de diversos fármacos [105]; [106].

As técnicas de modelagem molecular são realizadas mediante a combinação de diversos fatores estruturais e energéticos, relacionados principalmente à sua geometria estrutural e eletrônica, assim como à contribuição dos fatores dinâmicos e cinéticos energéticos. Em relação ao comportamento do composto molecular biologicamente, o mesmo é realizado mediante os aspectos tridimensionais [107].

A docagem molecular, também denominada ancoragem ou atracamento, é uma técnica *in silico* indispensável para diversos estudos quimioinformáticos. Seu potencial é possibilitar as estimativas energéticas das interações por meio da ligação entre proteína e ligante, antes que ocorra a síntese destes compostos [108]. Por intermédio da ancoragem molecular, são possíveis as obtenções de diversas conformações espaciais do ligante, fornecendo o prognóstico do modo da conformação mais suscetível para interagirem com o composto estrutural proteico (alvo). A obtenção das energias livres de ligação pode ser analisada com base em cada conformação apresentada entre os ligantes e o alvo, dessa forma é possível diagnosticar que a explicação mais provável para o modo conformacional da interação são as energias menores [109]. As conformações e orientações (poses) dos ligantes são viabilizadas por meio de algoritmos de busca, e a classificação ocorre através dos sítios de ligação em afinidade a seus termos, por meio das funções score ou a denominada pontuação [110]; [111]; [112].

Complexos moleculares nos quais ocorrem interações entre fármaco-receptor, ou seja, ligante e alvo proteico, acontecem por intermédio das forças intermoleculares em substâncias que possuem o tipo de interação denominada dipolo-dipolo induzido e permanente. As ligações de hidrogênio são recorrentes nas forças dipolo-dipolo permanentes, ou simplesmente dipolo-dipolo, enquanto nas forças de Van Der Waals ou London as interações do tipo dipolo-dipolo induzido são as mais comuns [113].

A ancoragem molecular, ou simplesmente docking, objetiva principalmente a realização do estudo por meio do cálculo da energia livre de ligação não covalente, realizada através de simulações entre ligante e proteína. Seu esquema representativo é representado pelas interações como Proteína mais Ligante e Ligante ↔ Proteína. O cálculo da energia livre do complexo proteína e ligante é executado mediante a subtração da energia livre do complexo ΔG_{LP}

(lectina-peroxidase), ΔGL (lectina) e ΔGP (peroxidase), em que a combinação de equações resultará na fórmula expressa como: $\Delta GLigação = GProteína-ligante - GProteína - Gligante$. Essa equação possibilita a obtenção da energia livre de todo o complexo enzimático após o cálculo de docagem molecular [114].

Para mensurar a energia livre de ligação e realizar a avaliação correspondente ao complexo receptor e ligante, existem diversas funcionalidades realizadas por meio de metodologias computacionais de docagem molecular. O processo de execução da simulação de ancoragem molecular deve seguir alguns parâmetros, dentre eles considerar a flexibilidade do ligante por meio dos procedimentos de método randômico. Por meio do uso de algoritmos, os ligantes são submetidos a mudanças randômicas. Outro método essencial para esse tratamento é o sistemático, com a finalidade de realizar a análise dos graus de liberdade do composto molecular [115]. A docagem molecular é o terceiro método essencial para o tratamento da flexibilidade do ligante, pois esse método possibilita a simulação com ligante e proteína simultaneamente flexíveis, ligante como uma entidade rígida e proteína inteiramente flexível, além de simulações com ligante e proteína rígidos [116]; [117].

3.8 Dinâmica Molecular

Historicamente, as simulações de dinâmica molecular (DM) ocorreram em meados dos anos 50, por Alder e T. E. Wainwright, com o objetivo de análise das interações entre duras esferas, visando principalmente o conhecimento do 'paradoxo da reversibilidade' [227]; [228]. Um grande avanço nos estudos relacionados à dinâmica molecular de um sistema real ocorreu em 1974, quando os pesquisadores Rahman e Stillinger realizaram a simulação de água líquida através de um potencial realístico [229].

O método de dinâmica molecular possibilita a simulação computacional versátil, que visa o estudo de macromoléculas biológicas através dos movimentos físicos de átomos e compostos moleculares, tomando como base a mecânica clássica newtoniana, por meio de informações sobre o comportamento mecânico que dispõe de informações sobre o campo de força ou potencial interatômico. Isso permite entender e compreender o comportamento dinâmico no quesito estrutura e movimentos dos átomos [230]. Dessa forma, é possível prever por meio dessas simulações como as biomoléculas irão responder em relação ao nível atômico a perturbações como mutação, fosforilação, protonação ou a adição ou remoção de um ligante. Ambos podem ser investigados mediante simulações de DM [231]. Nos estudos que envolvem planejamento, síntese e desenvolvimento de medicamentos, essas simulações de complexos no

nível atômico são mecanismos eficientes na caracterização dos comportamentos envolvendo as interações entre proteínas e ligantes, assim como as interações abrangidas na sinalização celular para diversas doenças [232].

De maneira geral, as simulações de DM possibilitam uma diversidade de informações energéticas relacionadas a proteínas, ligantes e interações moleculares, e não apenas a compostos estruturais dinâmicos e informações relacionadas a biomacromoléculas [233]. Os detalhes atômicos pertencentes a cada estrutura conformacional presente na biologia molecular são fornecidos mediante a utilização do método de DM na engenharia de proteínas [234].

Além disso, essa ferramenta metodológica de simulações possui diversas contribuições, dentre elas a capacidade de fornecer detalhes dos mecanismos que podem se agregar após o dobramento das proteínas, por meio da ligação química ou automontagem, visando compreender mecanismos. Contribuições no sentido de melhoramento referente ao aperfeiçoamento da triagem virtual mediante as diversas conformações do alvo no decorrer do processo de encaixe, possibilitando uma melhor e mais eficiente classificação de moléculas que apresentam êxito quando executadas, associado com cálculos de energia livre de ligação, além de possibilitar a descoberta de métodos eficientes de resistência a múltiplas drogas [235].

Os métodos de dinâmica molecular contribuíram significativamente nos estudos de docagem molecular e triagem virtual, pois ambos consideram a proteína flexível. Esse fato está relacionado principalmente ao reconhecimento da proteína pelo ligante, que ocorre através da natureza dinâmica, o que acarreta no rearranjo dos compostos estruturais. Tais ajustes na disposição estrutural integram o processo de ligação e energético. De maneira geral, as simulações de DM são essenciais para aumentar a precisão e eficiência dos resultados de docagem molecular [236].

Em relação ao campo de força em métodos de dinâmicas moleculares, são definidos como o conjunto dos potenciais de interações entre partículas. A execução dos cálculos em uma simulação computacional de DM é desenvolvida mediante o uso de um modelo de campo de força da mecânica molecular, que possibilita adequar resultados de cálculos da mecânica quântica. Tipicamente, nessas simulações, um campo de força engloba termos que realizam captura de interações eletrostáticas (atrações coulombianas) entre átomos [237].

Na atualidade, durante o processo de simulações computacionais de DM, os campos de força mais frequentemente empregados são HARvard (CHARMM) [238]. Desse modo, essas abordagens de simulação possibilitam avaliar inúmeras interações de partículas, mediante bilhões de etapas de tempo, que necessitam de diversos hardwares computacionais e tempo para sua execução [239]. O amplo potencial de aplicação desse método é realizado mediante a

utilização de diversos softwares que executam essas funções de cálculos de DM, dentre eles o GROMACS [221], CHARMM [240], dentre outros.

4 OBJETIVOS

4.1 Objetivo Geral

O presente estudo preconizou a realizar de uma triagem *in silico*, por meio de diversas combinações e criações diversificadas através da docagem molecular, baseada na aplicação de descritores moleculares e em sistema de aprendizado de máquina embasados em banco de dados que reúnem informações sobre produtos naturais, visando obter compostos promissores com atividade biológica contra doença de Chagas, utilizando um banco de dados de alcaloides isolados da família Euphorbiaceae.

4.2 Objetivos Especificos

- Construir um banco de dados de alcaloides isolados da família Euphorbiaceae, visando a elaboração de um perfil químico;
- Elaborar um modelo de predição no KNIME, com compostos com potencial atividade contra *T cruzi* através do banco de dados da família Euphorbiaceae;
- Realizar docagem molecular para diversos alvos da doença de Chagas;
- Analisar as interações moleculares existente através da docagem molecular, visando auxiliar nos estudos de triagem virtual e/ou relação estrutura-atividade;
- Investigar a relação estrutura-atividade e docagem molecular de derivados enzimas objetivando o planejamento de novos antiparasitários;
- Analisar a toxicidade dos alcaloides estudado mais promissores, visando a predição dos compostos potencialmente mais ativos, para condução futura de testes biológicos.
- Estudar através de simulações de dinâmica molecular a flexibilidade de interações dos melhores resultados de complexos proteína-ligante.

5 METODOLOGIA

5.1 Banco de Dados

A elaboração do banco de dados foi iniciada por meio do levantamento bibliográfico em busca de artigos relacionados a alcaloides isolados de gêneros e espécies pertencentes à família Euphorbiaceae. As palavras-chave utilizadas na pesquisa foram "Euphorbiaceae" e "Alkaloids". A prospecção dos artigos ocorreu na base de dados Web of Science, abrangendo um total de 254 artigos publicados entre os anos de 1967 e 2022. Em seguida, foi realizado o desenho e a otimização das geometrias, além da análise conformacional. Dessa forma, foi elaborado um banco de dados contendo quase 400 compostos pertencentes exclusivamente à classe de metabolitos secundários alcaloides, provenientes de diversas espécies de Euphorbiaceae, visando a obtenção dos códigos SMILES.

A notação SMILES (Simplified Molecular Input Line Entry Specification) é amplamente utilizada para denotar moléculas e representar graficamente estruturas químicas. Essa ferramenta possui uma compreensão simples e é largamente empregada no suporte de pesquisas e testes em softwares e bibliotecas da química computacional. Os SMILES podem representar estruturas moleculares por meio de caracteres (notação linear), representando o grafo da molécula e a fórmula esquelética. Sua descrição é única, visando evitar a repetição de códigos para moléculas estruturalmente diferentes. Dessa forma, ao adicionar esses códigos nos softwares e bancos de dados, ocorre automaticamente a conversão para informações sobre a estrutura do composto analisado. Isso possibilita a comparação com outros compostos moleculares presentes nos bancos de dados, visando a obtenção de alvos moleculares promissores a partir dos dados de similaridade [182].

O banco de dados de estruturas químicas ChEMBL (<https://www.ebi.ac.uk/chembl>) foi utilizado para a construção do conjunto de estruturas químicas visando a elaboração do modelo de predição. O conjunto apresenta 1707 estruturas químicas que foram previamente estudadas por meio de investigações *in vitro* para inibir a forma evolutiva tripomastigota da doença de Chagas. As estruturas foram classificadas, assim como as amastigotas, a partir do valor de PIC50. Dessa maneira, foi viável classificar 912 compostos como ativos (PIC50: 9,0-5,0) e 795 como inativos (PIC50: 4,99-1,75), utilizando como parâmetro de corte o valor de 5,0. É importante ressaltar que o valor de IC50 representa a concentração necessária para uma inibição de 50%.

Para as estruturas, foram utilizados códigos SMILES como dados de entrada para o software Marvin Sketch, da ChemAxon. Além disso, foi utilizado o programa Standardizer (JChem; ChemAxon). Esse programa computacional realiza a conversão de compostos químicos em representações canônicas personalizadas. Através da padronização, é possível

criar bibliotecas de compostos consistentes, além de obter as estruturas em formato canônico. O programa também realiza simultaneamente a inserção de hidrogênios, a aromatização, a geração do composto na geometria 3D e, por fim, salva as estruturas em formato SDF (Structure Data Format). Na quimioinformática, as estruturas químicas são essenciais para a execução de testes. No entanto, a manipulação e a ausência de preparação inadequadas podem resultar em falhas nas análises. Uma inspeção estrutural correta pode aumentar a garantia da representação precisa dos dados estruturais.

As planilhas de um banco de dados catalogadas no ChEMBL necessitam de organização e tratamento de dados. Isso ocorre porque os diversos procedimentos de organização asseguram que as linhas e colunas da planilha estejam corretamente alinhadas em relação a cada molécula exclusivamente. Dessa forma, ocorre a remoção das moléculas duplicadas presentes em cada planilha. Nesse caso, o procedimento foi realizado através do Microsoft Excel 2016. No entanto, pode ser realizado através de outros programas de computador, como o HIT QSAR, por exemplo. Outro procedimento essencial para o uso eficiente desses bancos de dados é a padronização, que se refere à conversão das unidades químicas. Essa abordagem possibilita uma melhor organização dos conjuntos de dados dos compostos, principalmente em relação às propriedades químicas semelhantes.

Nos estudos *in silico*, a utilização de planilhas baixadas de bancos de dados ou até mesmo aquelas construídas a partir de compostos isolados é essencial para padronizar e remover duplicatas dos compostos, possibilitando uma análise quimioinformática eficaz e confiável. Isso ocorre porque resultados provenientes de estudos computacionais exigem uma entrada de dados de excelente qualidade nos softwares, a fim de obter saídas com dados de confiança.

5.2 Descritores Volsurf

Os descritores moleculares nesse estudo, foram gerados através do software Volsurf, + v.1.0.7 visando prever a existência de propriedades biológica e físico-química do conjunto de compostos moleculares de banco de dados catalogados no ChEMBL. A realização do cálculo dos descritores, possibilita a transformação dos compostos moleculares em representação molecular que possibilita procedimentos matemáticos com os conjuntos de dados gerados nesses programas [183]; [184].

O programa computacional Volsurf permite a realização aproximadamente 128 descritores moleculares, para aplicação do cálculo dos campos de interação molecular

(*Molecular Interaction Fields*, MIFs) e GRIND (GRid-INdependent Descriptors), visando principalmente no processo de caracterização da ligação dos ligantes a seus receptores. O campo de interação MIF é fundamentado na energia total das interações moleculares entre alvos e um sonda química (molécula) e realiza o mapeamento das forças (ligações químicas), presente entre os compostos [184]; [185].

VolSurf + objetiva na produção e exploração no campo das propriedades físico-química de um composto molecular a qual se estar em processo de estudos, podendo ser explorado também bibliotecas virtuais moleculares, utilizando como base diversos mapas de energia de interação no formato 3D. Esse programa possui a finalidade de compactação dos mapas 3D para modelos de atividades de triagem virtual de simples compreensão, baseado em descritores numéricos otimizados para modelos de descritores ADME (absorção, distribuição, metabolismo e excreção), uma ferramenta relevante que otimizam estruturas químicas simultaneamente seguindo critérios que alta eficiência [184].

De maneira geral, o software VolSurf fornece aos pesquisadores ferramentas de grande relevância para o desenvolvimento de estudos, com a finalidade de possibilitar uma interpretação química mais precisa das matrizes de descritores moleculares. Dentre as principais ferramentas apresentadas nesse programas recebem destaque a Análise de Componentes Principais (PCA, “Principal Component Analysis”), Consenso PCA (CPCA, Consensus Principal Component Analysis) e Métodos Mínimos Quadrados Parciais (PLS, Partial Least Squares), entendidas como técnicas de análises quimiométricas que são métodos multivariáveis que pode ser utilizada na caracterização de uma rede complexa que conecta várias entidades biologicamente relevantes (sistema biológico) [184]; [186].

No VolSurf o cálculo de descritores ocorre através da importação do arquivo contendo os compostos moleculares salvo no formato especial de arquivo de dados (SDF), que corresponde a um arquivo de entrada.

5.3 Modelo de Predição

5.3.1 KNIME Analytics

O modelo de predição da atividade biológica neste estudo foi executado utilizando o programa computacional KNIME (versão 4.6.4), também conhecido como Konstanz Information Miner, com direitos autorais em www.knime.org. Esse software possibilita a manipulação, análise e modelagem de diversos dados por meio de programação visual, além da criação de modelos *in silico* por meio de um mapa auto-organizável através de "nós", nos quais

os dados são inseridos para análise. O conjunto dessas características forma um fluxo de trabalho (workflow) denominado de modelo de predição [187]; [188].

A plataforma KNIME Analytics proporciona uma maneira mais dinâmica por meio do aprendizado de máquina, uma subárea da inteligência artificial baseada na premissa de que a máquina (computador) pode aprender com os dados (modelo) a realizar a classificação de cada composto químico. O objetivo é realizar a predição da atividade, utilizando como base as características presentes em cada estrutura em estudo. Nesse sentido, com base na comparação das características das estruturas, ocorre a seleção de moléculas com base na similaridade. Durante a classificação do modelo preditivo, essas moléculas recebem classificação como ativas. As características dos compostos moleculares ativos e inativos, gerados por meio do aprendizado de máquina, possibilitam que a máquina, utilizando o método da comparação, classifique as moléculas para as quais a predição foi realizada. Dessa forma, essa ferramenta pode realizar a classificação de compostos que possuam atividade biológica não relatada na literatura [189]; [190].

A análise de similaridade é baseada nos descritores moleculares de compostos moleculares que possuam maior atividade, bem como nos compostos moleculares com atividade biológica desconhecida [191]. No caso deste estudo, a análise comparativa foi realizada com moléculas naturais. Posteriormente, as moléculas com atividade empregadas na elaboração do modelo preditivo foram divididas em uma planilha do Excel. Para esta divisão, os compostos moleculares ativos aplicados na elaboração do modelo foram divididos ao meio, utilizando como parâmetro o valor de pIC_{50} ($-\log(IC_{50})$). Compostos que apresentavam valores mais elevados foram designados como ativos (A), enquanto aqueles com valores mais baixos foram considerados inativos (I). Dessa forma, a máquina executa a classificação de diversos outros compostos estruturais que se enquadram no domínio de aplicação do modelo preditivo, conforme a metodologia descrita nesta pesquisa.

5.3.2 *Random Forest*

Neste estudo, o algoritmo *Random Forest* (Floresta Aleatória em português) foi utilizado na predição, considerando que esse algoritmo é empregado em aprendizado de máquina, conferindo maior confiabilidade e precisão ao estudo [192]; [193]. De maneira geral, esse algoritmo de aprendizado permite a criação de diversas árvores de decisão (do inglês Decision Trees), realizando uma investigação randômica das moléculas (em consenso). Dessa

forma, é possível combinar o conjunto de árvores de decisão para finalmente alcançar o resultado final [194]; [195].

O banco de compostos moleculares ativos contra *T. cruzi*, na forma tripomastigota, foram utilizados para a elaboração do modelo preditivo. Nesse sentido, foi realizada uma divisão de 80% para treino e 20% para teste. Dessa forma, o resultado final referente à classificação corresponde à ocorrência mais frequente em cada composto molecular [196]. Após a divisão do banco de dados em treino e teste, a máquina deu início ao processo de classificação dos compostos químicos com base na similaridade. As moléculas que, no início da execução dessa metodologia, foram classificadas como ativas, apresentaram um maior grau de similaridade com os compostos moleculares ativos. Respeitando o domínio de aplicação do modelo, maior será a probabilidade de que o composto estrutural em processo de predição apresente atividade, ou seja, seja ativo [197];[198].

Dessa forma, para que ocorra uma predição aceitável, o composto molecular precisa necessariamente estar inserido no domínio do modelo (aplicabilidade). Caso contrário, poderão ocorrer incorreções na predição, visto que pode não haver semelhanças ou mesmo compreensão integral. Em conformidade com esse sistema (modelo), que é baseado na análise de similaridade por meio de um processo comparativo, o domínio de aplicabilidade assegura que o composto em processo de análise preditiva contenha características relacionadas à estrutura representada no amplo espaço químico dos compostos empregados na elaboração do modelo preditivo (compostos ativos contra as formas infectantes do *T. cruzi*), conforme analisado nesses estudos. Nesse contexto, quando os compostos testados estão em sintonia com os do modelo no mesmo espaço químico, a fidedignidade do modelo de predição é ampliada, e quanto mais elevado o índice percentual de aplicabilidade, mais confiável é a predição para os compostos moleculares estudados [199]; [200]; [201].

5.3.3 *Wecka Predictor 3.7*

No desenvolvimento deste estudo, foi utilizada a versão 3.7 da biblioteca *Wecka Predictor*, com o objetivo de realizar a classificação de compostos moleculares que possuem atividade anti-*T. cruzi* desconhecida pela literatura [202]; [203]. Este método, associado ao Random Forest, corresponde ao aprendizado de máquina. Além disso, tanto a classificação da extensão *Wecka Predictor* quanto o modelo de predição (mapa auto-organizável) elaborado pelo software KNIME, criado no workflow, estão cada vez mais consolidados em pesquisas

envolvendo aprendizado de máquina, principalmente devido à sua confiabilidade e eficiência preditiva [204]; [205].

Em relação aos resultados do modelo, a avaliação da confiabilidade baseou-se na estrutura de uma matriz de confusão, que é caracteristicamente uma tabela que apresenta os erros e acertos presentes no modelo preditivo em comparação com os resultados esperados. Os dados que garantem a confiabilidade do modelo de predição são a avaliação da acurácia, cujo índice deve ser de 70% ou mais, assim como o coeficiente de correlação de Matthews (MCC), que deve atingir 50% ou mais [206]. A avaliação do poder discriminatório da análise do desempenho do modelo de predição neste estudo foi realizada por meio da análise do estimador estatístico curva ROC (*Receiver Operating Characteristic*), tanto nos testes quanto na validação interna de desempenho e no domínio de aplicação [207].

5.3.4 Curva ROC

Nesta pesquisa, foi realizada a avaliação do modelo de predição por meio da análise da curva ROC, uma ferramenta essencial para verificar o desempenho do modelo preditivo. Durante o processo de análise, observa-se o gráfico bidimensional que apresenta o desempenho executado pelo modelo preditivo, abrangendo todas as classificações disponíveis na máquina. Dessa forma, foi possível avaliar, por meio de uma abordagem comparativa, o desempenho dos modelos classificatórios. Isso foi feito mediante a definição dos critérios classificadores para bons e maus pagadores, baseados na seleção dos pontos de corte em toda a amplitude determinada pelos escores constatados pelo modelo preditivo executado [208]; [209].

O resultado referente à performance da representação gráfica da curva ROC, oriunda do modelo de predição, tem como base a sensibilidade (fração dos verdadeiros positivos - VP) que se refere aos ativos, assim como a especificidade (taxa de verdadeiros negativos - VN) que se refere aos inativos, além da precisão relacionada à previsibilidade geral com base nos valores do teste. Além disso, foram utilizadas a sensibilidade (verdadeiro positivo) na variável dependente ou eixo y, assim como a especificidade (verdadeiro negativo) na variável independente ou eixo x. Dessa forma, a interpretação da curva ROC deve ocorrer analisando o percurso (traçado) da curva, pois quanto maior a distância da curva em relação à diagonal principal, mais satisfatório será o desempenho do modelo avaliado com base na curva [210]; [211].

Complementarmente, foi aplicado neste estudo o cálculo da área abaixo da curva ROC (AUC), do inglês "*Area Under the ROC Curve*". Essa ferramenta metodológica é amplamente utilizada em estudos de análise de modelos de predição, uma vez que possui um domínio

estatístico relevante. Seu objetivo é medir e analisar de maneira comparativa o desempenho de modelos de classificação. De maneira geral, a AUC corresponde à área entre a curva ROC e o eixo X, com variação de 0 a 1. O parâmetro utilizado para avaliar os resultados dessa comparação de desempenho são os valores adquiridos durante os estudos. Uma AUC com valor igual a 1,0 ou mais próximo indica que o teste possui a máxima capacidade de classificação, ou seja, apresenta seu melhor desempenho. Enquanto a AUC igual a 0,5 sugere que o teste está seguindo a linha diagonal do gráfico, sendo interpretada como uma performance ruim, ou seja, o pior desempenho do teste [210]; [212].

O fluxo do aprendizado de máquina (trabalho) referente ao modelo preditivo executado nessa pesquisa encontra-se disponível na figura 6.

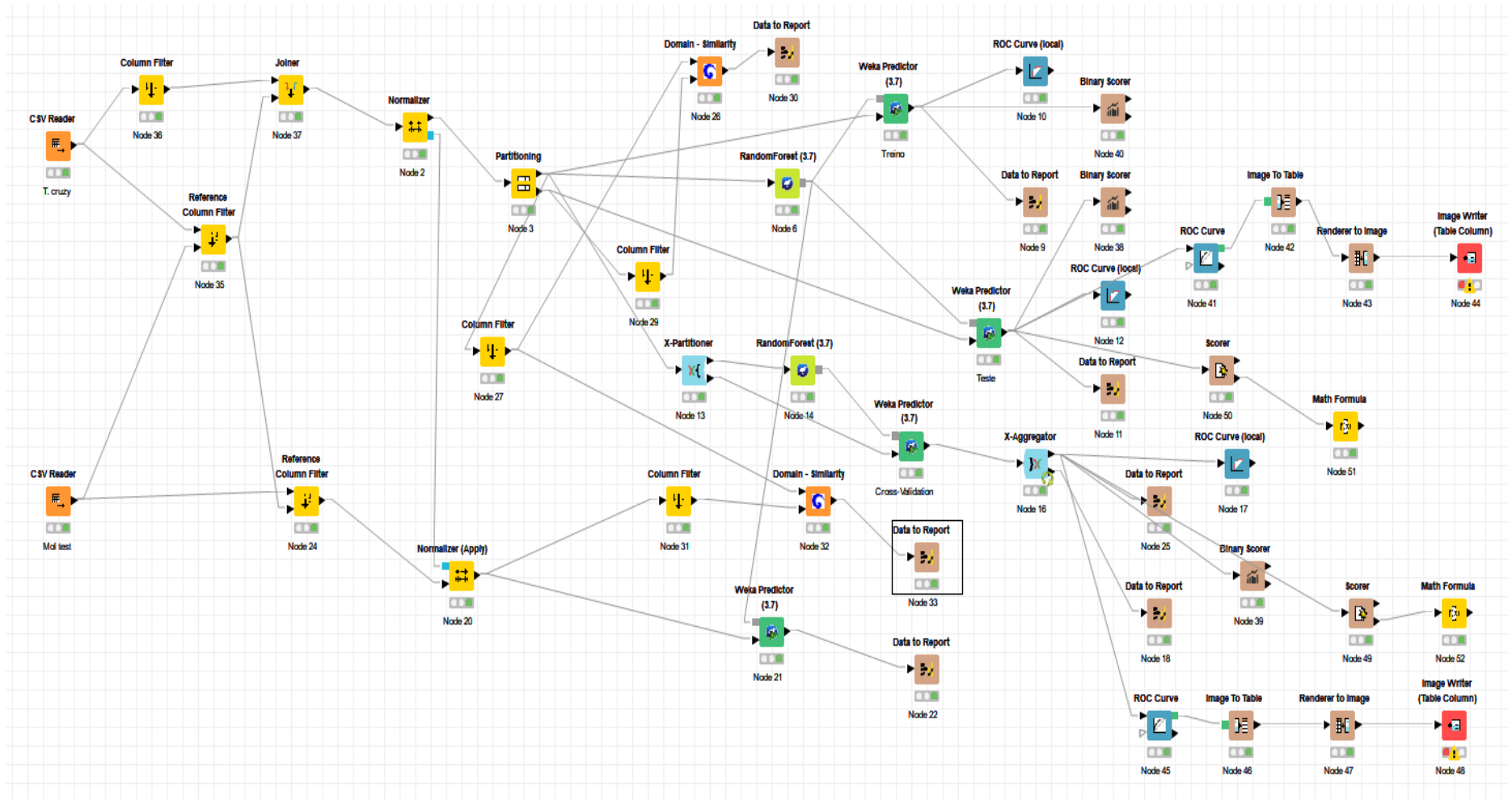


Figura 6. Fluxo de trabalho (Workflow) empregado na construção do modelo e realização das análises estatísticas
 Fonte: Próprio autor, 2023.

5.4 Docagem Molecular

Para a ancoragem molecular com o objetivo de fixar as moléculas, foi utilizado neste estudo o programa computacional Molegro Virtual Docker 6.0.1 (MVD), um software proprietário. As proteínas-alvo do *T. cruzi*: 1GXF [213]; 1MXF [214]; 3HBB [215]; 3LXS [216], juntamente com seus ligantes respectivos, foram baixadas do site Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/home/home.do>). Os principais dados das enzimas estão apresentados na Tabela 1. Para a execução do acoplamento molecular neste software, as moléculas de alcaloides isolados da família Euphorbiaceae foram importadas, juntamente com as proteínas-alvo mencionadas anteriormente. Inicialmente, foram removidas as moléculas de água e cofatores. Em seguida, foi realizada a criação do *template*, tendo como base os inibidores complexados com as proteínas-alvo estudadas neste trabalho. Além disso, foram utilizados outros parâmetros predefinidos pelo Molegro para a execução da ancoragem [217].

Além do procedimento de acoplamento molecular realizado por meio do *docking*, em seguida foi executado o redocking, com o objetivo de obter o valor referente ao RMSD (*Root Mean Square Deviation*, em inglês). Esse cálculo dos desvios quadráticos médios é obtido por meio da análise do desvio quadrático médio entre os átomos do ligante complexado juntamente com a proteína e seus respectivos ligantes, com base nos mesmos parâmetros do *docking* inicial. Dessa forma, é possível validar o protocolo utilizado para o acoplamento molecular. Os valores de RMSD dos ligantes obtidos por meio do redocking são considerados válidos para *docking* que apresentam valores abaixo de 3.0 Å, os quais indicam uma precisão ideal no acoplamento molecular [218]; [219].

Tabela 1. Informações das enzimas alvos para *T. cruzi* selecionadas, depositadas no banco de dados PDB e usadas para ancoragem.

Enzima PDB ID	Nome	Ligand PDB	Resolution
1GXF	Tripanotiona redutase	Quinacrine Mustard	2.70 Å
1MXF	Pteridine Reductase 2	Methotrexate	2.30 Å
3HBB	Dihydrofolate reductase-thymidylate synthase (DHFR)	Trimetrexate	3.00 Å
3LXS	Cruzaína	Inibidor Piperazine	1.50 Å

Fonte: Próprio autor, 2023.

Através da aplicação dos métodos apresentados para a docagem molecular, é possível estabelecer as seguintes interpretações dos resultados do processo referente ao *docking* realizados neste estudo. É possível a obtenção das energias e análise dos tipos de interações. Nesse caso, o software utilizado apresenta três tipos de interações: hidrogênio (representadas por linhas tracejadas azuis), eletrostáticas (representadas por linhas tracejadas verdes) e estéricas (representadas por linhas tracejadas vermelhas). Além disso, no processo de interpretação dos resultados, é possível observar os resíduos de aminoácidos envolvidos em cada interação entre os ligantes e os respectivos receptores. Através desse procedimento, é possível encontrar resíduos comuns entre os compostos análogos e controles, e até mesmo inibidores complexados em conjuntos com as proteínas estudadas. O objetivo é correlacionar a relevância das interações que possam ocorrer para a atividade biológica.

5.5 Previsão De Propriedades ADMET

Foram realizados cálculos dos parâmetros de absorção, distribuição, metabolismo e excreção (ADME), utilizando a ferramenta web de acesso aberto SwissADME (<http://www.swissadme.ch>). Esta ferramenta oferece um conjunto de modelos preditivos rápidos para a avaliação de propriedades físico-químicas, farmacocinéticas e farmacológicas, entre outras. Quanto à previsão da toxicidade, utilizou-se o software OSIRIS Property Explorer (<https://www.organic-chemistry.org/prog/peo/>), um explorador de diversas propriedades de toxicidade presentes em drogas, compostos sintéticos, entre outros. Além disso, essa ferramenta realiza a avaliação da absorção considerando múltiplos fatores, como permeabilidade da membrana, absorção intestinal e se o composto recebe a classificação de substrato ou inibidor da enzima glicoproteína de permeabilidade P-glicoproteína (P-gp), além de destacar se o inibidor foi considerado [225]; [226].

Neste estudo, foram investigados compostos moleculares que não ultrapassaram no máximo duas violações da regra de Lipinski e apresentaram um consenso logP de até 4,15. Além disso, os substratos da glicoproteína de permeabilidade (P-gp) não foram eliminados ou excluídos da análise. A distribuição dos compostos foi avaliada considerando diversos fatores, incluindo a barreira hematoencefálica (logBB) e a permeabilidade do sistema nervoso central (SNC). Quanto à análise do metabolismo, as previsões foram realizadas por meio de modelos para substratos ou inibidores das enzimas do citocromo P450 (CYP1A2, CYP2C19, CYP2C9, CYP2D6 e CYP3A4). A toxicidade foi outro parâmetro avaliado, sendo sua previsão realizada

por meio da mutagenicidade, tumorigenicidade, tumorigenicidade, efeitos colaterais na reprodução e irritabilidade.

5.6 Simulações de Dinâmica Molecular

Simulações de Dinâmica Molecular foram realizadas para estimar a flexibilidade das interações entre proteínas e ligantes, usando o software GROMACS 5.0 (European Union Horizon 2020 Program, Suécia) [220,221]. A topologia dos ligantes foi preparada usando o Automated Topology Builder (ATB) e Repository Versão 3.0 (<https://atb.uq.edu.au/>) [222] aplicando o campo de força GROMOS96 5a7. A simulação MD foi realizada usando o modelo de água SPC de carga pontual, estendido em uma caixa cúbica [223]. O sistema foi neutralizado pela adição de íons (Cl⁻ e Na⁺) e minimizado, para remover maus contatos entre moléculas complexas e o solvente. O sistema também foi balanceado a 300K, usando o algoritmo V-rescale de 100 ps, representado por NVT (número constante de partículas, volume e temperatura), seguido de equilíbrio à pressão de 1 atm, usando o algoritmo de Parrinello-Rahman como o NPT (Constant Partícula de pressão e temperatura), até 100 ps. Simulações DM foram realizadas em 5.000.000 passos, a 10 ns. Para determinar a flexibilidade da estrutura e se o complexo é estável próximo à estrutura experimental, os valores de RMSD de todos os átomos de C α foram calculados em relação às estruturas iniciais. Os valores de RMSF também foram analisados para entender os papéis desempenhados pelos resíduos próximos ao sítio de ligação do receptor. Os gráficos RMSD e RMSF foram gerados no software Grace (Grace Development Team, <http://plasma-gate.weizmann.ac.il/Grace/>) [224,264].

6 RESULTADOS E DISCUSSÃO

6.1 Análise do Banco De Dados

O banco de dados foi construído inicialmente mediante busca eletrônica de artigos na base de dados Web of Science (<http://apps-webofknowledge.ez15.periodicos.capes.gov.br/>), com a finalidade de delimitar a busca pelos artigos foi usado como palavra chave na busca “Euphorbiaceae” and “Alkaloids”. Dessa forma, foram catalogados artigos relacionados a alcaloides isolados de gêneros e espécies pertencentes a família Euphorbiaceae. Nessa pesquisa foram encontrados 254 artigos no total, publicados entre os anos de 1967 e 2022, os artigos formam utilizados para a construção do banco de dados de alcaloides isolados da família Euphorbiaceae.

Através desse método, o banco de dados foi construído e compreendeu quase 400 compostos, sendo 314 estruturas diferentes, pertencentes a 73 espécies. Esses compostos naturais são alcaloides que compõem o banco construído, pertencente a diversos gêneros e espécies da família Euphorbiaceae. Dentre o total de moléculas presentes no banco de dados, os principais gêneros dos compostos naturais catalogados pertencem ao gênero *Croton* (80), seguidos por *Acalypha* (70), *Euphorbia* (46) e *Jatropha* (41), além de outros que estão representados por cores na figura 7. Os compostos moleculares que fazem parte do banco de dados, contendo aproximadamente 400 compostos, encontram-se no anexo I, juntamente com seus respectivos códigos SMILES.

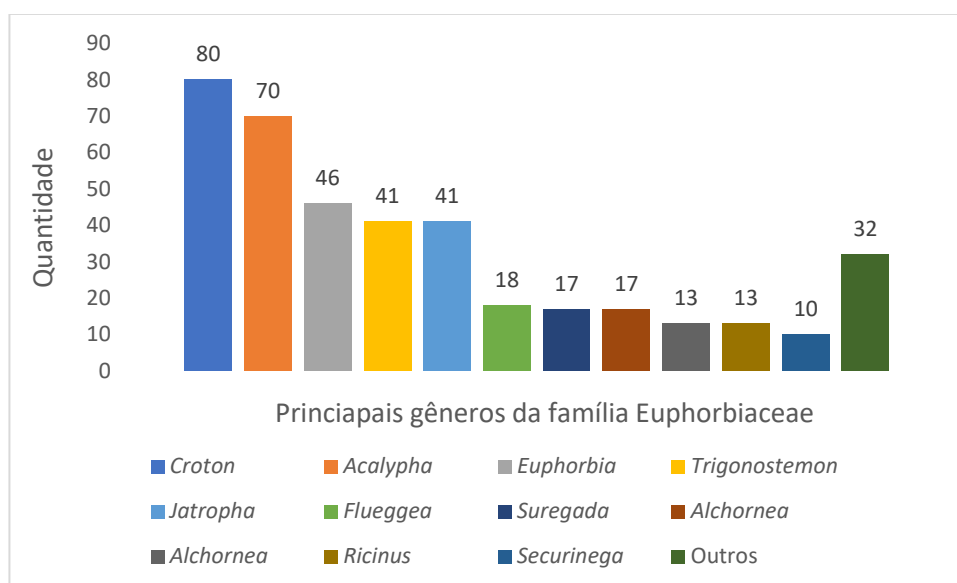


Figura 7. Principais gêneros dos alcaloides de Euphorbiaceae presente no banco de dados.
Fonte: Próprio autor, 2023.

6.2 Análise dos Modelos de *T. Cruzi* na Forma Tripomastigosta

Inicialmente, todas as moléculas estudadas, tanto os compostos ativos quanto os candidatos, foram submetidas a uma série de tratamentos em software utilizados na quimioinformática. Isso visou a otimização e geração de descritores e banco de dados. Além disso, foram realizados tratamentos utilizando planilhas Excel, de forma a possibilitar a submissão nos modelos de predição.

Os SMILES das estruturas moleculares foram inseridos como dados de entrada no programa computacional VolSurf. Através desse software, foram gerados descritores para 1707 moléculas que possuem atividade conhecida contra a forma evolutiva tripomastigota (estruturas químicas disponíveis no ChEMBL). Em seguida, os descritores foram utilizados como dados

de entrada no programa computacional Knime, a fim de gerar o modelo preditivo. Dessa forma, foi realizada a construção de um modelo de predição utilizando descritores moleculares calculados pelo programa computacional VolSurf.

Para garantir uma proporção de previsões corretas pelo modelo, é essencial que o mesmo disponha de uma excelente confiabilidade preditiva. Com esse propósito, os dados utilizados no modelo de previsão foram analisados estatisticamente para confirmar a confiabilidade do modelo em estudo. Essa confirmação foi obtida através da avaliação dos valores estatísticos gerados pelo modelo de previsão executado no VolSurf.

Os descritores moleculares, juntamente com os dados referentes à atividade biológica qualitativa, foram empregados com o objetivo de gerar o modelo de predição utilizando o algoritmo '*Random Forest*' (RF), um método de aprendizado de máquina. Dessa forma, para o modelo de tripomastigota, os compostos foram classificados com base na pIC50 ($-\log IC_{50}$ em mol/l). Isso permitiu a divisão em grupos com atividade e inatividade, correspondendo a 912 compostos ativos (pIC50: 9,0-5,0) e 795 compostos inativos (pIC50: 4,99-1,75), utilizando um valor de corte de 5,0 como parâmetro.

Com o objetivo de aprimorar a classificação do modelo elaborado, uma margem de 0,5 foi adicionada, estabelecendo um intervalo para os valores de pIC50 dos compostos entre 5,3 e 4,79. Como resultado, 291 compostos foram excluídos, deixando 1417 compostos no conjunto de dados. Nesse contexto, os compostos considerados ativos totalizaram 751, sendo caracterizados por possuírem $pIC_{50} \geq 5,3-9,0$. Por outro lado, os compostos inativos foram identificados por possuírem valores de $pIC_{50} \leq 4,79-1,75$, totalizando 666 compostos.

6.2.1 Descritores VolSurf

O modelo preditivo para a forma tripomastigota da *T. Cruzi*, durante a análise, demonstrou, tanto na validação cruzada (cross validação) quanto no teste, desempenhos próximos em termos estatísticos. As taxas de acerto foram de aproximadamente 77%. O treinamento apresentou um desempenho ideal, atingindo uma taxa de acerto de 100%. Os resultados dos parâmetros obtidos no modelo preditivo são apresentados na Tabela 2.

Tabela 2. Resultados dos parâmetros estatísticos de confiabilidade do modelo de predição com compostos químicos de atividade conhecida para forma tripomastigota utilizando os descritores VolSurf.

Indicadores de confiabilidade do modelo preditivo							
Validação	Especificidade	Sensibilidade	Acurácia	PPV	NPV	MCC	AUC
Test	0,749	0,823	0,788	0,791	0,75	0,574	0,859
Cross	0,767	0,775	0,771	0,787	0,79	0,541	0,853

Legendas: MCC: coeficiente de correlação de matthews; ACC: acurácia; AUC: área sob a curva; PPV: valores preditivos positivos; NPV: valores preditivos negativos.

Fonte: Próprio autor, 2023.

A determinação dos valores percentuais da sensibilidade e especificidade do modelo de predição foi realizada por meio de análise estatística. A sensibilidade corresponde à proporção de verdadeiros positivos (VP) em relação ao total de casos positivos reais (VP + FN). Em outras palavras, ela mensura a capacidade do modelo em identificar corretamente os casos positivos. A fórmula da sensibilidade é a seguinte: $\text{Sensibilidade} = \text{VP} / (\text{VP} + \text{FN})$. A especificidade é a proporção de verdadeiros negativos (VN) em relação ao total de casos negativos reais (VN + FP). Ela avalia a capacidade do modelo em identificar corretamente os casos negativos. A fórmula da especificidade é: $\text{Especificidade} = \text{VN} / (\text{VN} + \text{FP})$ [262]. Neste estudo, os modelos de predição executados para a forma tripomastigota apresentaram bons níveis percentuais nos parâmetros de especificidade e sensibilidade. As medidas estatísticas que avaliaram o desempenho do modelo de classificação alcançaram 76% em relação às taxas de ativos e inativos para a série de testes. No que diz respeito à validação cruzada, o percentual foi de 74%.

Estes dados possibilitaram a realização dos cálculos para o Coeficiente de Correlação de Matthews–MCC, viabilizando a avaliação do modelo na forma tripomastigota, que possibilita a correlação das classificações binárias observadas e preditivas, em que o MCC varia de -1 a +1, onde +1 representa uma previsão perfeita, 0 corresponde um desempenho aleatório e -1 indica uma previsão totalmente incorreta, ou seja, uma incompatibilidade entre a predição e observação realizada no modelo preditivo. Nesse estudo, na forma tripomastigota, obteve-se para o grupo de teste um MCC correspondente a 0,57, e para validação cruzada, valores de 0,54. Dessa forma, os valores obtidos no modelo indicam que ele possui uma boa capacidade de predição. Isso se deve ao fato de que valores próximos a +1 indicam uma alta concordância entre as previsões e os resultados reais. Em outras palavras, apresenta valores que se aproximam de uma predição perfeita. As medidas referentes ao MCC são essenciais para avaliar a eficácia e confiabilidade do modelo, permitindo uma análise mais precisa do comportamento da detecção de casos positivos e negativos.

Adicionalmente, foi gerado o gráfico da Curva Característica de Operação do Receptor (ROC), visando analisar o desempenho do modelo de predição desenvolvido. A área sob a curva gerada para a forma tripomastigota no conjunto de teste foi de 0,859; já para a validação cruzada, esse valor foi de 0,853 (conforme ilustrado na Figura 8).

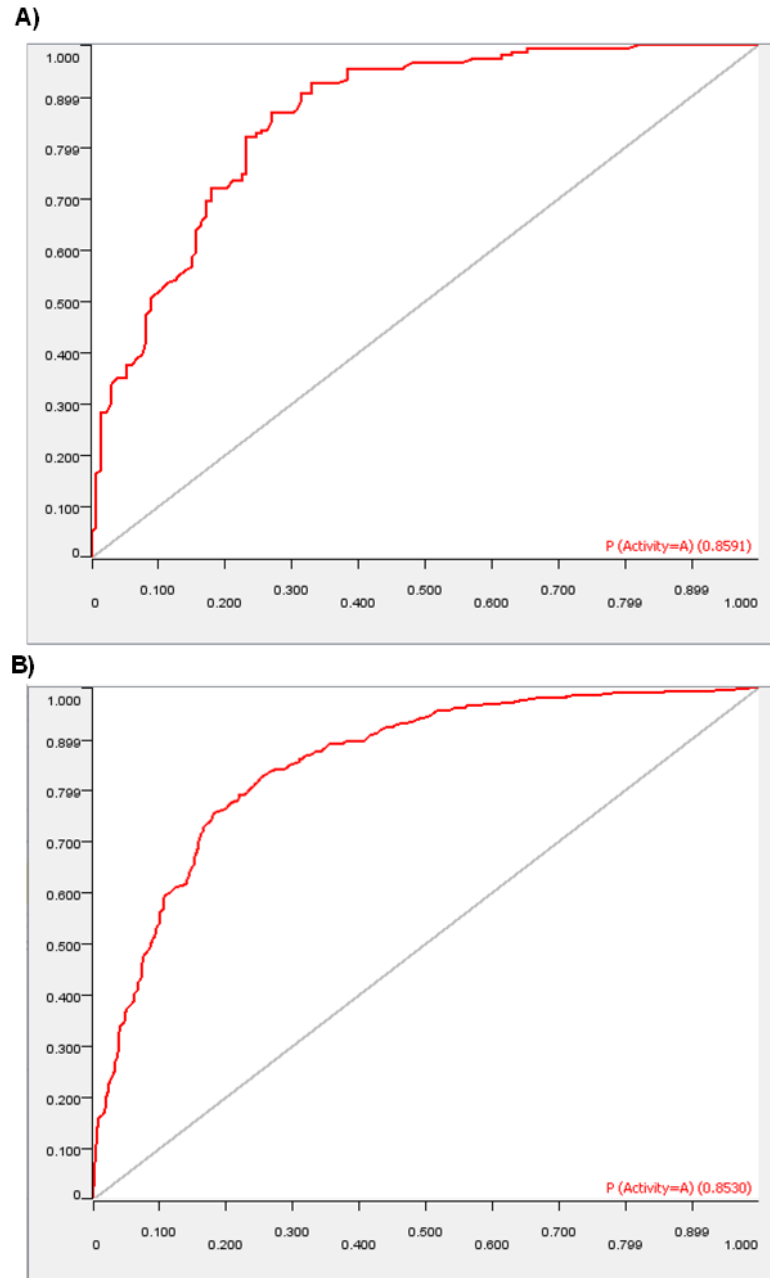


Figura 8. Gráfico ROC com área sob a curva para o conjunto de teste do modelo de *T cruzi* obtido com “Random Forest”.

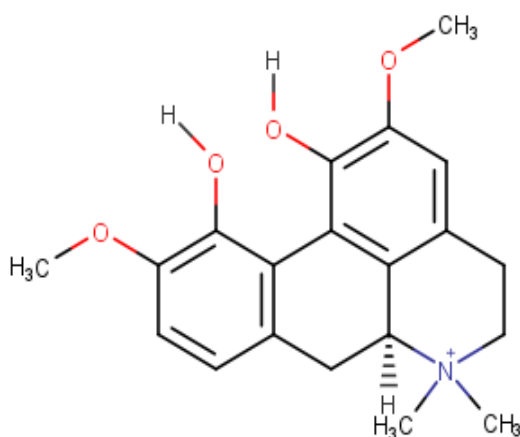
Legendas: Forma tripomastigota – a) Teste; b) Validação Cruzada.

Fonte: Próprio autor, 2023.

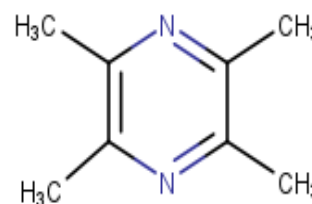
A partir dos resultados apresentados acima, é possível afirmar que o modelo de predição executado neste estudo possui uma alta taxa de classificação. Isso ocorre porque um modelo perfeito possui uma área sob a curva igual ou próxima de 1, como demonstrado no gráfico ROC. Dessa forma, pode-se inferir que este modelo tem um desempenho aceitável na tarefa de classificação.

O modelo desenvolvido para a forma tripomastigota classificou cerca de 91 moléculas com potencial atividade (ativas); ou seja, foi atribuída ao modelo uma probabilidade de atividade para cada molécula, sendo que aquelas com uma probabilidade acima de 50% são consideradas potencialmente ativas. Dentre as moléculas selecionadas, 14 obtiveram atividades com probabilidade na faixa de 75% a 80%; 19 compostos obtiveram percentuais de atividade com valores equivalentes de 65% a 74%; 35 compostos apresentaram percentuais de atividade na faixa de 55% a 64%; e um outro conjunto de compostos, os 23 restantes, apresentou probabilidade com valores entre 50% a 54%. A eficácia do método foi inspecionada através do domínio de aplicabilidade, o qual demonstrou boa confiabilidade para todos os compostos moleculares classificados como ativos.

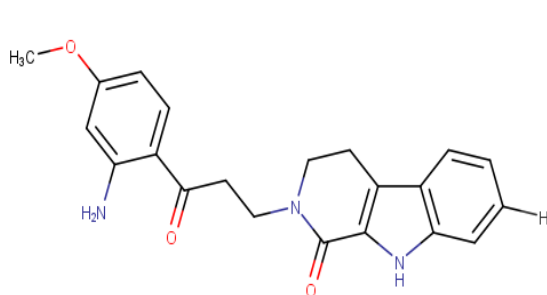
Com o objetivo de ampliar a seleção, optou-se por filtrar apenas as moléculas com uma probabilidade de atividade acima de 70%, totalizando 21 moléculas selecionadas. Os compostos moleculares que possuem o maior percentual de probabilidade de atividade estão apresentados na figura 9.



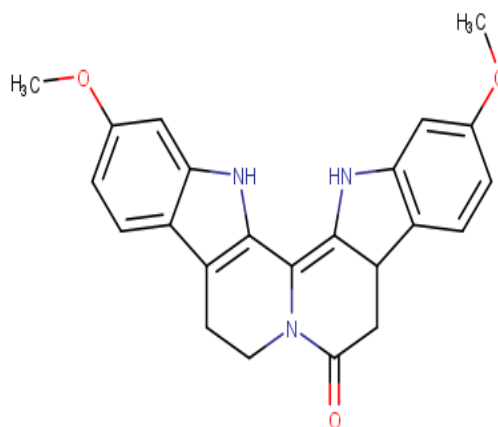
Composto 149 (Magnoflorine);
P= 88%



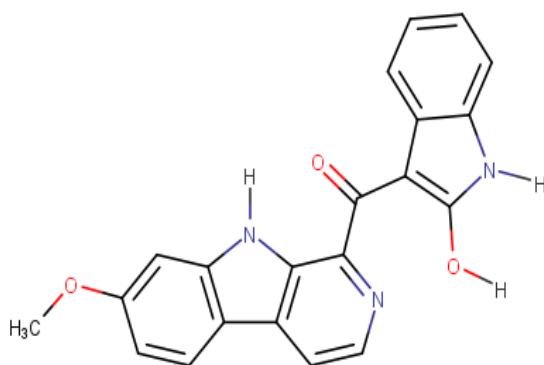
Composto 194 (Tetramethylpyrazine);
P= 86%



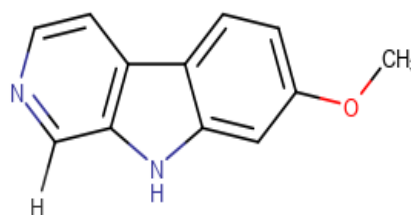
Composto 150 (Trigonostemonine A);
P= 86%



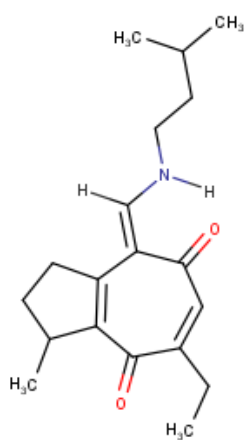
Composto 333 (trifiline C);
P= 83%



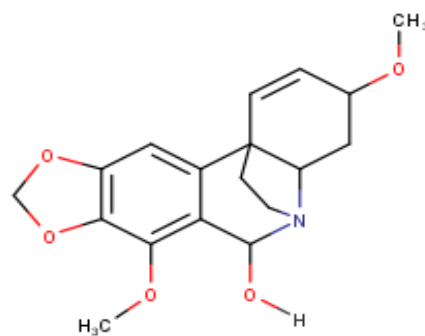
Composto 320 (Trigonostemine C);
P= 80%



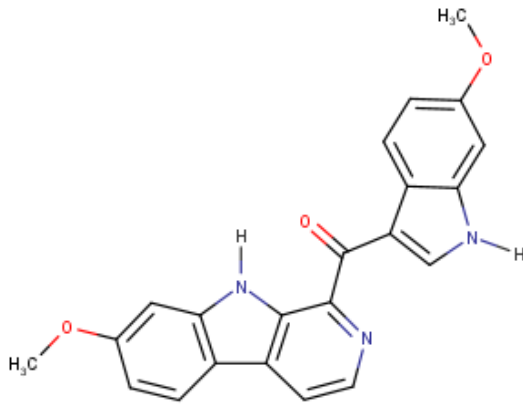
Composto 326 (1,2,3,4-tetrahydro-oxo-β-carboline);
P= 79%



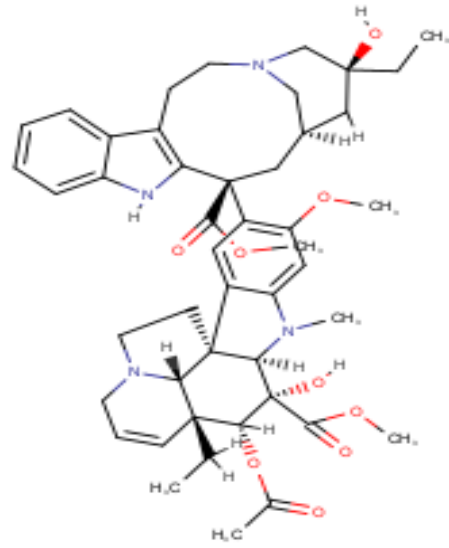
Composto 114 (Muscicapine A);
P= 78%



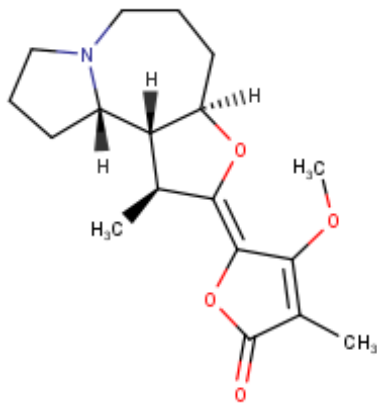
Composto 222 (6-Hydroxy-buphanidrine);
P= 77%



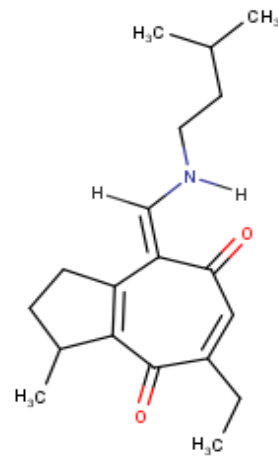
Composto 319 (Trigonostemine B);
P= 76%



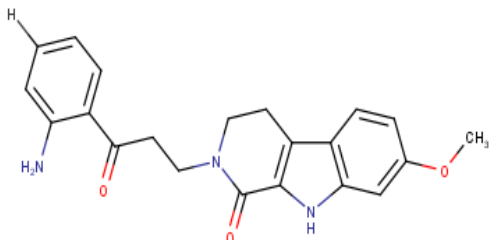
Composto 367 (Vinblastine);
P= 76%



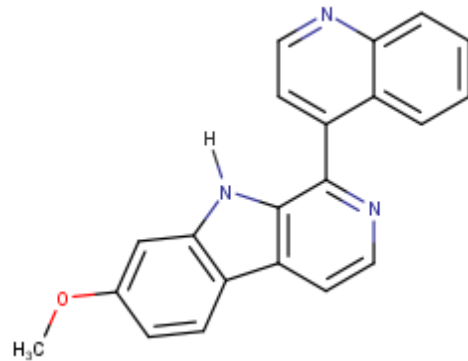
Composto 038 (Isoprotostemone);
P= 75%



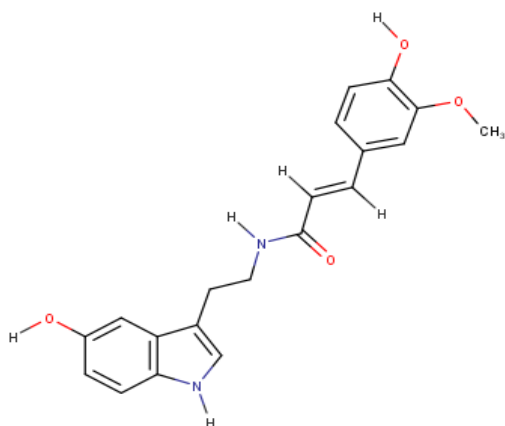
Composto 115 (muscapine B);
P= 74%



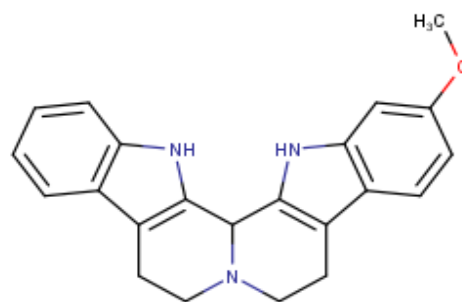
Composto 151 (Trigonostemonine B);
P= 74%



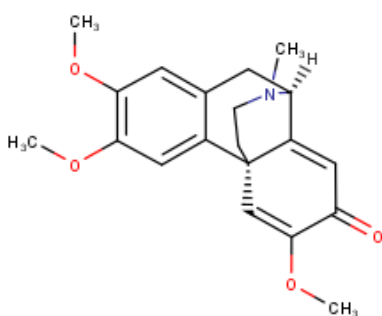
Composto 152 (Trigonostemonine C);
P= 74%



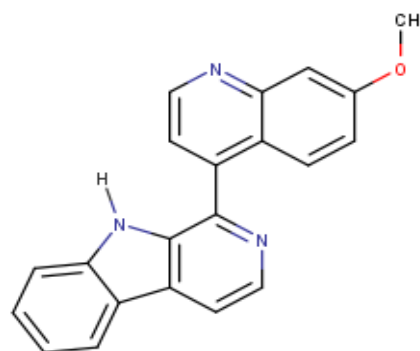
Composto 309 (Moschamine);
P= 74%



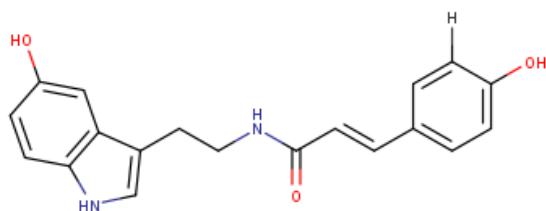
Composto 315 (Trigonostemonine F);
P= 73%



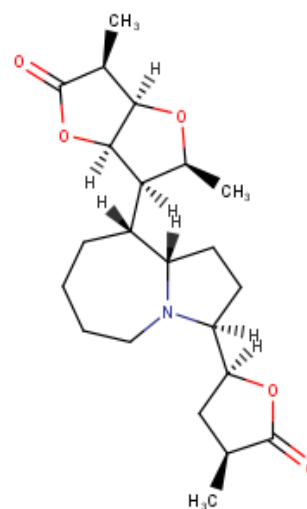
Composto 369 (O-Methylpallidine);
P= 73%



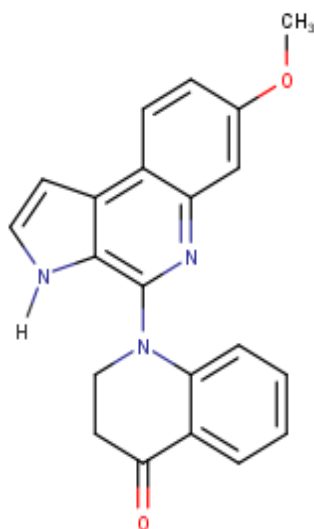
Composto 153 (Trigonostemonine D);
P= 73%



Composto 350 (N-trans-p-Coumaroyl-5-hydroxytryptamine);
P= 73%



Composto 002 (Parvistemonine);
P= 70%



Composto 329 (Trigonoine B);

P= 70%

Figura 9. Compostos moleculares (arquitetura 2D) em estudo com maiores percentuais de probabilidade de atividade para a forma tripomastigota de *T. cruzi*.

* Os nomes dos compostos estão disponíveis no anexo I.

Fonte: Próprio autor, 2023.

Dentre as moléculas apresentadas acima, outras também demonstraram bons percentuais referentes à probabilidade de potencial atividade, incluindo os compostos 316, 370 e 395, com índices de 69%. O composto 139 apresentou um percentual de 67%, o composto 22 demonstrou 66% de probabilidade de atividade, além dos compostos 393, que possui uma probabilidade de 65%, assim como os compostos 39, 55, 60, 70, 86, 103, 328, 334, 92, 125, 145, 245, 279, 305, 355, 381, 346, 202, 318, 327, 345 e 306, com potencial atividade alcançando entre 60 e 69%. Todos esses compostos podem ser selecionados para a realização de testes biológicos.

6.3 Docking Molecular

Para a realização da ancoragem molecular, foram selecionadas quatro proteínas-alvo contra o *T. cruzi*: 1GXF, 1MXF, 3HBB e 3LXS, todas extraídas do banco de dados PDB, juntamente com seus ligantes inibidores correspondentes: Quinacrina Mostarda, Metotrexato, Trimetrexato e Inibidor Piperazina.

Antes de realizar as simulações de docagem molecular, foram conduzidos testes de validação dos compostos em estudo por meio do redocking entre os ligantes e as proteínas co-cristalizadas. Foram utilizadas as seguintes estruturas cristalinas como referência: Tripanotiona

redutase para *T. cruzi* (PDB: 1GXF), em complexo com o Quinacrine Mustard como primeiro ligante; Pteridine reductase para *T. cruzi* (PDB: 1MXF), em complexo com o Methotrexate como segundo ligante; Dihydrofolate reductase-thymidylate synthase (DHFR) para *T. cruzi* (PDB: 3HBB), em complexo com o Trimetrexate como terceiro ligante; e a Cruzaína para *T. cruzi* (PDB: 3LXS), em complexo com o Inibidor piperazine como quarto ligante.

Os resultados do redocking para as enzimas Tripanotiona redutase (PDB: 1GXF), Pteridine reductase (PDB: 1MXF), Dihydrofolate reductase-thymidylate synthase (DHFR) (PDB: 3HBB) e a Cruzaína (PDB: 3LXS) estão apresentados na Tabela 3. Nela, encontramos os valores de energia obtidos por meio da pontuação realizada pela função de pontuação MolDock Score (KJ/mol) e o valor do RMSD (Desvio médio quadrático do ligante). O RMSD corresponde a uma medida que quantifica a diferença entre a posição do ligante na estrutura cristalina conhecida experimentalmente e a posição prevista durante a ancoragem. Quando o RMSD é inferior a 2.0 Å, isso indica que a predição da ancoragem molecular apresentou uma alta similaridade ou proximidade com a posição real do ligante na proteína.

Tabela 3. Valores referente as pontuações do MolDock para o ligante e RMSD.

Enzimas	Tripanotiona redutase (PDB: 1GXF)	Pteridine reductase (PDB: 1MXF)	DHFR (PDB: 3HBB)	Cruzaína (PDB: 3LXS)
Ligante	-82.2447	-47.8028	-82.1702	-143.814
Redocking	-77.1478	-52.5507	-80.1659	-130.958
RMSD	0.89	1.75	0.60	1.04

Legendas RMSD: Desvio médio quadrático do ligante; DHFR: Dihydrofolate reductase-thymidylate synthase. **Fonte:** Próprio autor, 2023.

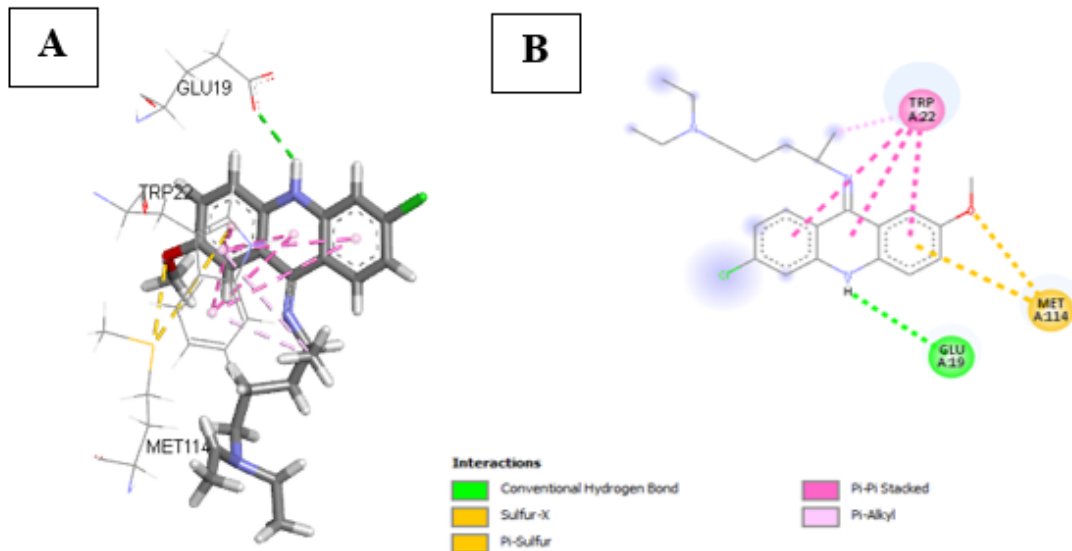
Os resultados obtidos indicam o sucesso e a validação do método, uma vez que os valores de RMSD foram de 0,89 Å para o quinino mostarda, ligante da proteína tripanotiona redutase (PDB: 1GXF); 1,75 Å para o metotrexato, ligante da proteína pteridina redutase (PDB: 1MXF); 0,60 Å para o trimetrexato, ligante da proteína DHFR (PDB: 3HBB) e 1,04 Å para o inibidor piperazina, ligante da proteína cruzaína (PDB: 3LXS). Ambos os valores estão abaixo de 2,0 Å, evidenciando uma excelente correspondência entre as posições preditas pelo docking e as posições reais dos ligantes nas proteínas. Além disso, observou-se que os valores de energia total foram baixos, o que indica uma forte interação e afinidade dos ligantes com as proteínas. Quanto menor a energia total, maior é a probabilidade de que a ligação entre o ligante e a proteína seja estável e favorável.

A visualização das interações estabelecidas foi conduzida utilizando o programa Visualizador do Discovery Studio, da BIOVIA, versão 2021 (<https://www.3dsbiovia.com/>). As ligações de hidrogênio estão destacadas em verde, as interações hidrofóbicas em rosa e as interações estéricas são realçadas em vermelho.

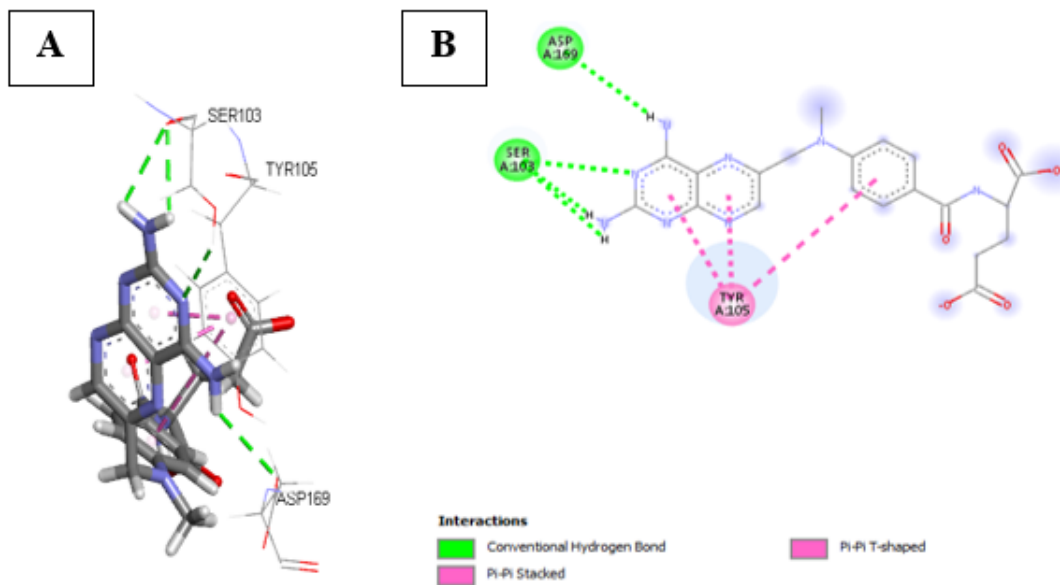
Os ligantes das proteínas em estudo estão localizados em seus respectivos sítios de ligação (uma região na superfície da proteína que acomoda o ligante), os quais são apresentados na figura 10, juntamente com as interações estabelecidas com os resíduos de aminoácidos das enzimas. Esses conjuntos de aminoácidos são responsáveis por permitir que o sítio ativo da enzima se encaixe exclusivamente em um alvo específico.

De maneira geral, é de fundamental importância compreender as conformações assumidas pelos ligantes no sítio ativo de uma proteína. Isso permite compreender a atividade e especificidade das interações moleculares envolvidas, viabilizando assim a projeção de inibidores e até mesmo a otimização das atividades de compostos em estudos envolvendo química medicinal.

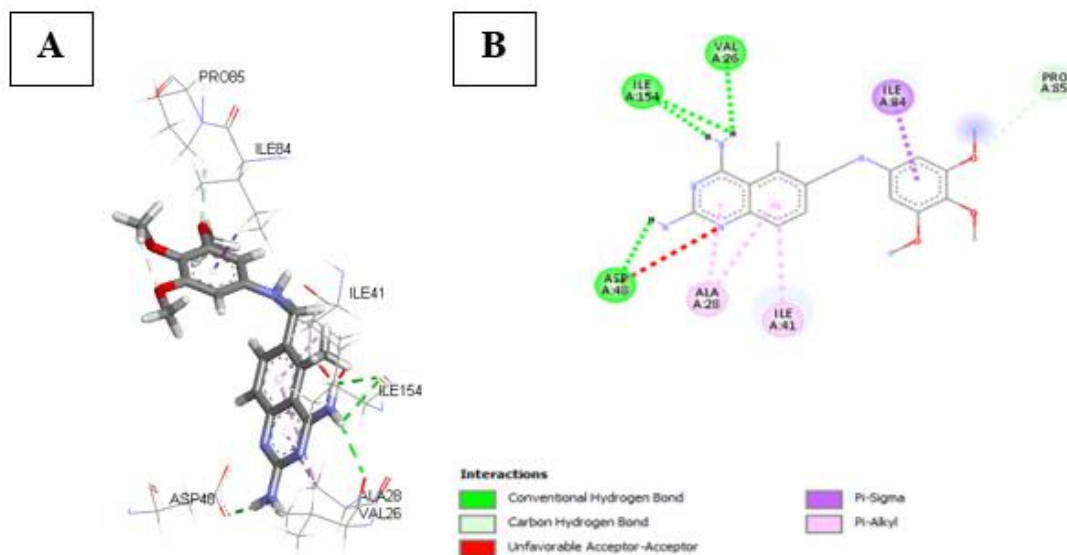
Enzimas: PDB: 1GXF



Enzimas: PDB: 1MXF



Enzimas: PDB: 3HBB



Enzimas: PDB: 3LXS

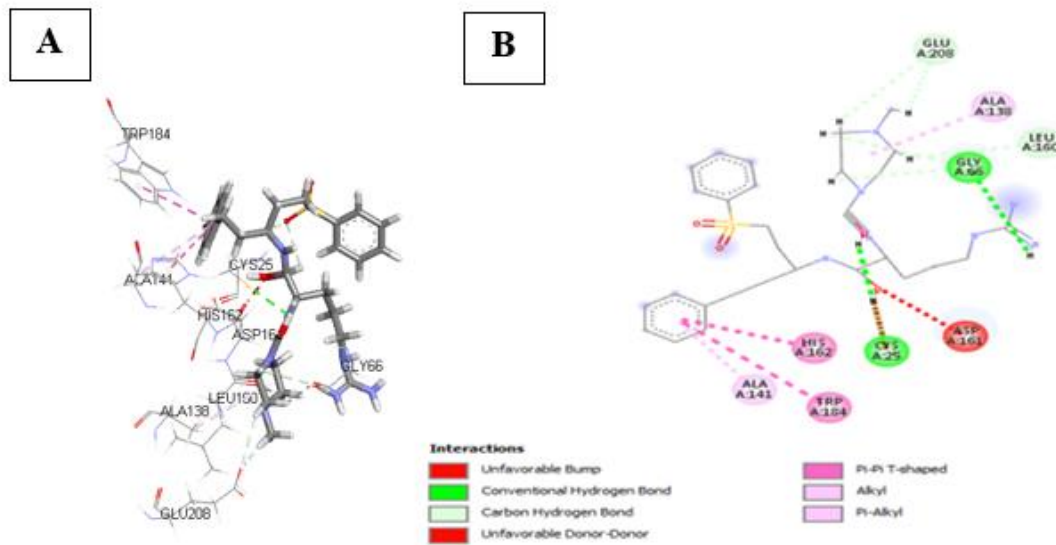


Figura 10. Ligantes e as interações correspondentes no sítio ativo das enzimas em 3D (A) e 2D(B).

Legenda. Interações (3D e 2D), as pontes de hidrogênio estão destacadas em verde, as interações hidrofóbicas são destacadas em rosa e as interações estéricas são destacadas em vermelho. ALA: Alanina; GLU: Ácido Glutâmico; ILE Isoleucina; LEU Leucina; MET Metionina; PHE Fenilalanina; TRP Triptofano; TYR Tirosina; VAL Valina; SER: Serina; CYS: cisteína; ILE: Isoleucina; ASP: Ác. Aspártico; HIS: Histidina.

Fonte: Próprio autor, 2023.

Uma coleção de 398 moléculas desenvolvidas a partir de alcaloides isolados de plantas da família Euphorbiaceae foi investigada nesta pesquisa. No entanto, dentre esse total, apenas 21 obtiveram um percentual de atividade de 70% no modelo de predição. Dessa forma, somente os 21 compostos foram submetidos a simulações de docking molecular. Esse conjunto de compostos foi submetido a testes utilizando as quatro proteínas-alvo de *T. cruzi*, com o propósito de identificar quais possuíam maior potencial de atividade contra o *T. cruzi*. A primeira enzima escolhida foi a de PDB 1GXF é a cruzipaina, que é uma enzima proteolítica presente no parasita *Trypanosoma cruzi*, o agente causador da doença de Chagas. Em diversos estudos *in silico* apresentou-se potenciais inibidores ou moléculas candidatas que possam interagir com a cruzipaina e inibir sua atividade. Isso é importante no contexto do desenvolvimento de novos medicamentos para o tratamento da doença de Chagas. A segunda enzima de código PDB 1MXF; 3HBB; 3LXS apresentou potenciais inibidores ou moléculas candidatas que possam interagir com o *T. Cruzi* em estudos *in silico*. Para essa análise, uma triagem virtual foi realizada, utilizando a estrutura do receptor como referência.

As energias de ligação e pontuações foram calculadas para as 21 moléculas do banco de dados, utilizando as funções de pontuação Moldock, PLANTS e Rerank. Em seguida, foram

realizados cálculos para identificar as moléculas com maior probabilidade de apresentar atividade potencial contra *T. cruzi*. Para cada função de pontuação, o valor "p" foi determinado, representando a relação entre a pontuação obtida para cada composto e a pontuação do composto que recebeu a menor energia, conforme indicado na equação a seguir.

$$\text{Prob} = \frac{E_{MT}}{E_M}, \text{ se } E_{MT} < E_L.$$

Nesta equação, comparamos a energia da molécula de teste (E_{MT}) com a menor energia obtida entre as moléculas testadas (E_M) e a energia do ligante derivado da cristalografia da proteína em teste (E_L). A fração é determinada pela condição de que a energia da molécula de teste seja inferior à energia do ligante. Em outras palavras, apenas as moléculas com energia de ligação menor do que a energia de ligação do ligante obtido pela cristalografia serão consideradas como potencialmente ativas.

Após obter os valores de probabilidade para os compostos em cada função de pontuação investigada, realiza-se o cálculo da probabilidade total (p). Esse valor é obtido somando as probabilidades de cada função de pontuação em estudo e dividindo pelo número total de observações, conforme demonstrado através da equação a seguir.

$$P_{\text{Enzyme}} = \frac{(\text{p})\text{MolDock Score} + (\text{p})\text{PLANTS Score} + (\text{p})\text{Rerank Score}}{n}$$

Foram utilizadas quatro proteínas distintas no conjunto de dados. Valores mais negativos indicaram melhores previsões em todas as funções de pontuação. Após a normalização dos resultados, calculou-se a média de todas as energias obtidas para cada composto em análise. As tabelas (4), (5), (6) e (7) a seguir apresentam os valores obtidos em cada pontuação do Moldock, PLANTS e Rerank, bem como a probabilidade total (p) dos compostos em relação a cada proteína estudada.

Tabela 4. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 1GXF.

ID	1GXF						
	Moldock	(p) Moldok	Rerank	(p) Rerank	PLANTS	(p) Plants	(p) Total
149	-111.811	1	-62.7561	0.127801378	-444.366	0.901479	0.676427
194	-49.7638	0.445071	-32.493	0.066171	-436.997	0.88653	0.465924
150	-105.482	0.943396	-60.9118	0.124046	-438.247	0.889065	0.652169

333	-70.4109	0.629731	-46.0474	0.093774	-489.091	0.992212	0.571906
320	-67.3662	0.602501	-42.6012	0.086756	-476.248	0.966157	0.551805
326	-71.4426	0.638959	-48.2331	0.098226	-491.677	0.997458	0.578214
114	-69.9312	0.625441	-52.1505	0.106203	-411.488	0.83478	0.522141
222	-39.8532	0.356434	40.1137	-0.08169	-492.93	1	0.424914
319	-57.1382	0.511025	-34.8665	0.071005	-480.068	0.973907	0.518646
367	-74.1985	0.663606	-43.0995	0.087771	-468.453	0.950344	0.56724
038	-93.7174	0.838177	-68.2465	0.138982	-478.798	0.971331	0.649497
115	-71.5184	0.639637	-56.8292	0.115731	-406.214	0.82408	0.526483
151	-49.851	0.445851	-29.7272	0.060539	-483.21	0.980281	0.495557
152	-50.7392	0.453794	-19.2203	0.039142	-476.586	0.966843	0.486593
309	-49.8157	0.445535	-491.044	1	-483.225	0.980312	0.808615
315	-68.4669	0.612345	-41.4241	0.084359	-471.239	0.955996	0.5509
369	-68.4082	0.61182	-42.6688	0.086894	-471.175	0.955866	0.551527
153	-65.6957	0.58756	-46.4641	0.094623	-477.161	0.96801	0.550064
350	-64.9022	0.580463	-46.6279	0.094957	-481.014	0.975826	0.550415
002	-96.1571	0.859997	-66.4276	0.135278	-438.555	0.88969	0.628322
329	-74.7361	0.668415	-43.9201	0.089442	-491.18	0.99645	0.584769
Ligante	-72.5897	0.649218	-42.0611	0.085656	-428.326	0.868939	0.534604

Fonte: Próprio autor, 2023.

Tabela 5. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 1MXF.

ID	1MXF						
	Moldock	(p) Moldok	Rerank	(p) Rerank	PLANTS	(p) Plants	(p) Total
149	-103.171	0.796509	-71.3019	0.824212	-403.51	1	0.873574
194	-68.5269	0.529047	-49.1564	0.568222	-342.373	0.848487	0.648585
150	-104.574	0.80734	-78.3131	0.905257	-383.869	0.951325	0.887974
333	-117.402	0.906376	-86.5092	1	-380.835	0.943806	0.950061
320	-71.9953	0.555824	-57.009	0.658993	-399.659	0.990456	0.735091
326	-86.052	0.664345	-70.0146	0.809331	-364.116	0.902372	0.792016
114	-71.0492	0.54852	-51.1905	0.591735	-262.992	0.651761	0.597338

222	-56.7106	0.437822	-46.4311	0.536719	-326.223	0.808463	0.594335
319	-71.9953	0.555824	-57.009	0.658993	-392.828	0.973527	0.729448
367	-100.727	0.777641	-66.9278	0.77365	-378.761	0.938666	0.829985
038	-129.529	1	-76.3544	0.882616	-370.326	0.917762	0.933459
115	-88.6243	0.684204	-64.3155	0.743453	-265.227	0.6573	0.694986
151	-104.574	0.80734	-78.3131	0.905257	-394.418	0.977468	0.896689
152	-75.4147	0.582223	-55.9474	0.646722	-400.19	0.991772	0.740239
309	-74.9394	0.578553	-52.885	0.611322	-394.416	0.977463	0.722446
315	-96.4408	0.74455	-66.8659	0.772934	-401.11	0.994052	0.837179
369	-96.4005	0.744239	-66.8302	0.772521	-401.301	0.994526	0.837095
153	-75.4147	0.582223	-55.9474	0.646722	-394.697	0.978159	0.735701
350	-73.8298	0.569987	-46.7998	0.540981	-392.773	0.973391	0.694786
002	-107.723	0.831652	-28.019	0.323885	-400.033	0.991383	0.71564
329	-92.0916	0.710973	-63.6422	0.73567	-393.117	0.974244	0.806962
Ligante	-112.99	0.872314	-86.2586	0.997103	-321.103	0.795775	0.888397

Fonte: Próprio autor, 2023.

Tabela 6. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 3HBB.

ID	3HBB						
	Moldock	(p) Moldok	Rerank	(p) Rerank	PLANTS	(p) Plants	(p) Total
149	-117.755	1	-93.592	1	-500.568	0.977643	0.992548
194	-83.0999	0.705702	-64.2222	0.686193	-443.552	0.866287	0.752727
150	-102.363	0.869288	-73.1552	0.781639	-483.585	0.944474	0.865134
333	-81.9805	0.696195	-16.6244	0.177626	-477.369	0.932334	0.602052
320	-91.391	0.776111	-69.8341	0.746155	-484.962	0.947164	0.823143
326	-88.3102	0.749949	-73.0453	0.780465	-466.516	0.911137	0.81385
114	-73.9543	0.628035	-46.0716	0.49226	-327.422	0.639477	0.586591
222	-63.8573	0.542289	-53.831	0.575167	-415.028	0.810578	0.642678
319	-90.4603	0.768208	-72.9332	0.779267	-493.201	0.963255	0.83691
367	-85.4271	0.725465	-40.4382	0.432069	-512.015	1	0.719178
038	-91.0535	0.773245	-65.3423	-0.69816	-449.095	0.877113	0.317399

115	-78.1535	0.663696	-61.5025	0.657134	-328.802	0.642173	0.654334
151	-102.363	0.869288	-73.1552	0.781639	-491.134	0.959218	0.870048
152	-84.7819	0.719986	-65.0095	0.694605	-488.16	0.95341	0.789333
309	-98.2227	0.834128	-73.5573	0.785936	-491.131	0.959212	0.859759
315	-98.9252	0.840093	-69.6689	0.744389	-500.795	0.978087	0.85419
369	-99.309	0.843353	-69.426	0.741794	-500.794	0.978085	0.854411
153	-80.2827	0.681777	-60.9953	0.651715	-489.02	0.955089	0.762861
350	-89.9375	0.763768	-67.6392	0.722703	-490.238	0.957468	0.814646
002	-117.268	0.995864	-83.9303	0.896768	-482.066	0.941508	0.944713
329	-97.8522	0.830981	-69.1875	0.739246	-480.685	0.93881	0.836346
Ligante	-79.6998	0.676827	-71.1042	0.759725	-393.863	0.769241	0.735265

Fonte: Próprio autor, 2023.

Tabela 7. Valores obtidos em cada pontuação Moldock PLANTS, Rerank com a proteína 3LXS.

ID	3LXS						
	Moldock	(p) Moldok	Rerank	(p) Rerank	PLANTS	(p) Plants	(p) Total
149	-106.533	0.929097	-30.6224	0.355076	-327.175	0.977587	0.75392
194	-53.8676	0.469791	-46.203	0.535739	-295.709	0.883568	0.629699
150	-103.134	0.899453	-28.167	0.326605	-334.059	0.998156	0.741405
333	-103.803	0.905288	-80.9881	0.939083	-314.612	0.940049	0.92814
320	-97.3998	0.849444	-60.7224	0.704096	-310.168	0.926771	0.82677
326	-85.8709	0.748898	-57.526	0.667032	-306.526	0.915889	0.777273
114	-103.574	0.903291	-86.2417	1	-231.992	0.693184	0.865491
222	-73.4069	0.640197	-60.204	0.698085	-284.04	0.848701	0.728994
319	-93.3703	0.814302	-65.7255	0.762108	-313.163	0.93572	0.837377
367	-114.663	1	-72.6372	0.842251	-334.676	1	0.947417
038	-96.6765	0.843136	-8.39083	0.097294	-301.948	0.90221	0.614213
115	-94.6207	0.825207	-77.9179	0.903483	-232.687	0.69526	0.807983
151	-87.3188	0.761526	-63.7905	0.739671	-310.585	0.928017	0.809738
152	-99.0415	0.863762	-69.2693	0.8032	-306.199	0.914912	0.860624
309	-95.8814	0.836202	-69.3821	0.804508	-310.581	0.928005	0.856238

315	-109.757	0.957214	-64.0253	0.742394	-330.729	0.988207	0.895938
369	-77.0827	0.672254	-64.5552	0.748538	-331.598	0.990803	0.803865
153	-96.597	0.842443	-53.5204	0.620586	-306.313	0.915252	0.79276
350	-74.9495	0.65365	-7.37864	0.085558	-313.129	0.935618	0.558275
002	-105.089	0.916503	-77.2559	0.895807	-304.027	0.908422	0.906911
329	-95.5374	0.833202	-68.312	0.792099	-321.497	0.960622	0.861974
Ligante	-144.263	0.794819	-91.8764	0.938671	-504.407	0.663504	0.798998

Fonte: Próprio autor, 2023.

As energias de ligação obtidas através do Moldock são representadas por valores negativos. Ao aplicar essa fração na equação, os valores são transformados em uma escala de 0 a 1, permitindo que os valores negativos sejam interpretados como positivos.

As substâncias analisadas exibiram interação com as quatro enzimas avaliadas, uma vez que as energias obtidas foram negativas. Além disso, constatou-se que em todos os quatro mecanismos investigados, algumas moléculas apresentaram energias mais baixas do que o ligante PDB, o que indica uma interação e afinidade significativas.

Dentre as 21 moléculas analisadas, foram identificadas 18 com potencial de atividade contra a enzima 1GXF. Todas as 21 moléculas demonstraram potencial de atividade contra a enzima 1MXF. As 21 moléculas apresentaram potencial de atividade contra a enzima 3HBB e, por fim, todas as 21 moléculas exibiram potencial de atividade contra a enzima 3LXS. Os compostos moleculares que mostraram energias mais baixas que os ligantes em estudo em cada enzima analisada estão apresentados na tabela 8.

Tabela 8. Valores referente a pontuação do Moldock para o ligante e das moléculas que obtiveram menores energias e maiores valores de probabilidade nas funções de pontuação Moldock, PLANTS e Rerank.

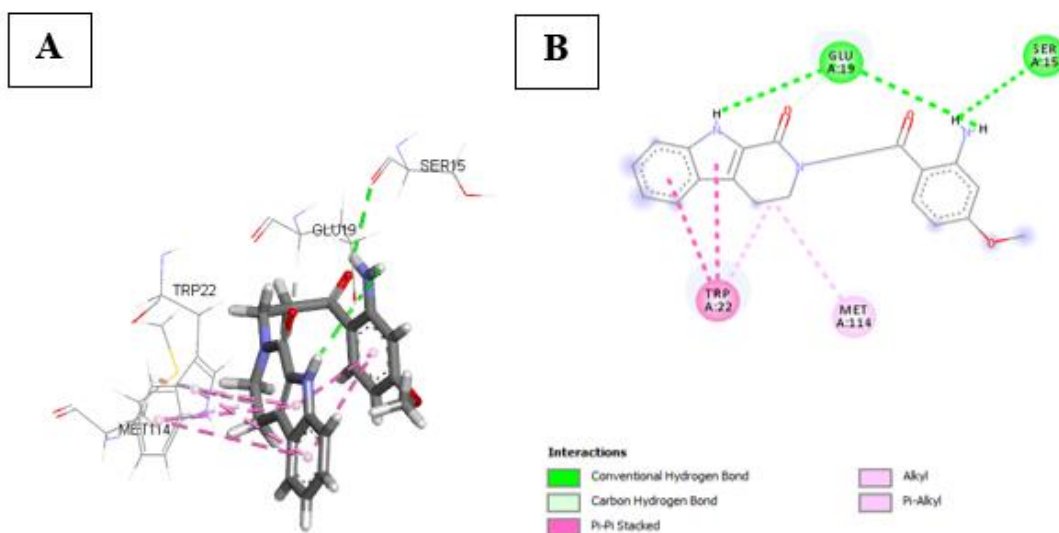
PROTEINAS	1GXF	1MXF	3HBB	3LXS
Energia Ligante	-72.5897	-112.99	-79.6998	-144.263
Moléculas com menores energias	-111.811 (Composto 149)	-129.529 (Composto 038)	-117.755 (Composto 149)	-114.663 (Composto 367)

* Os nomes dos compostos estão no anexo I. A proteína 3LXS não apresentou composto com energia mais baixa que o ligante, entretanto foi adicionado na tabela o composto 367 pois foi o que possuiu melhor energia em relação ao ligante.

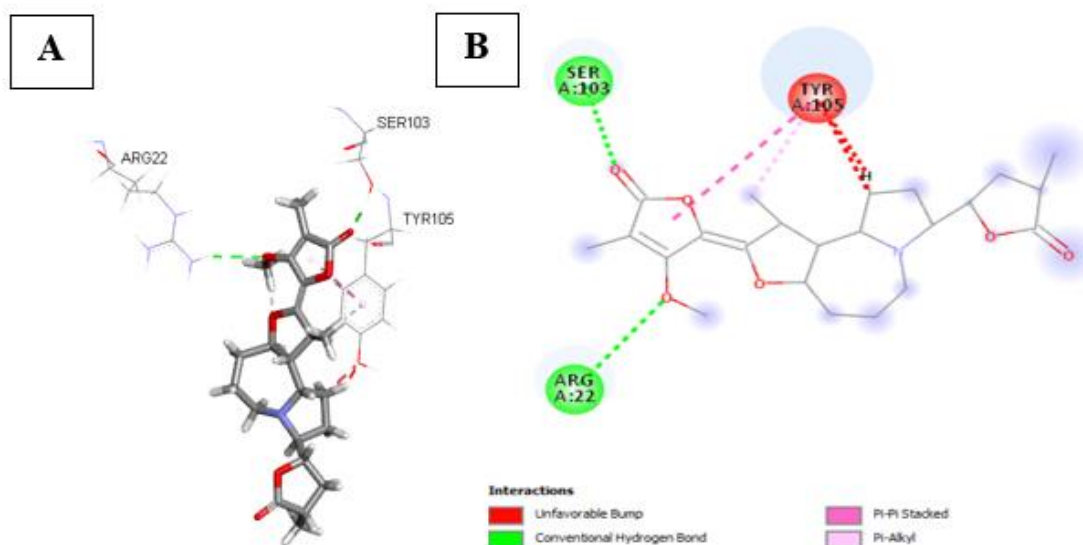
Fonte: Próprio autor, 2023.

Análises e discussões foram conduzidas acerca dos resíduos envolvidos nas interações entre os ligantes e os receptores dos compostos com as melhores energias. Para esta análise, foram considerados apenas os resíduos de uma única molécula, aquela com a menor energia, em cada proteína investigada. As interações e os resíduos correspondentes estão representados na Figura 11 a seguir.

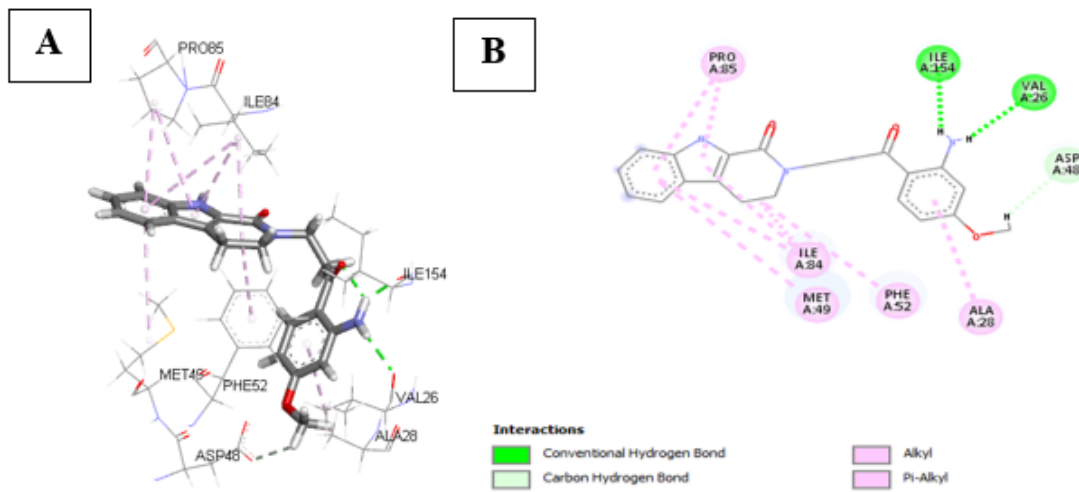
1GXF



1MXF



3HBB



3LXS

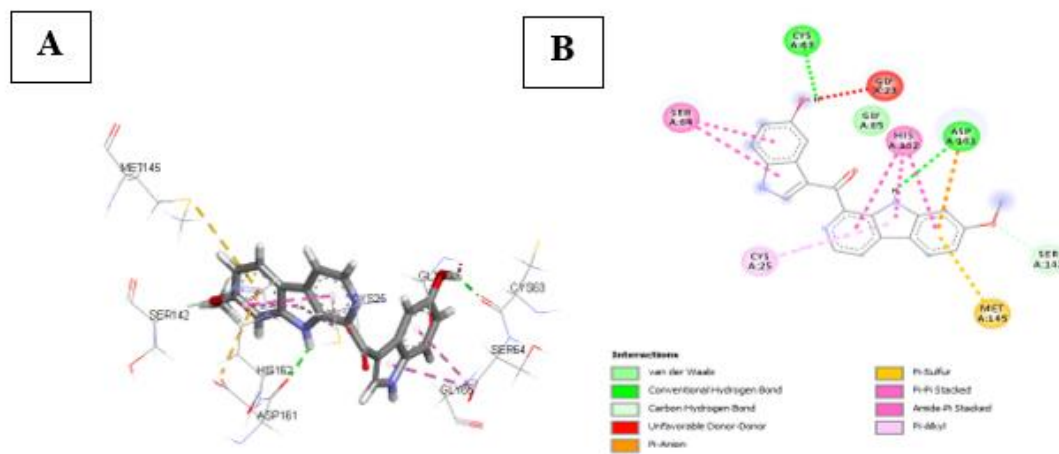


Figura 11. Interações 3D (A) e 2D (B) de alcaloides com os melhores valores de energia de ancoragem frente a *T. cruzi*.

Fonte: Próprio autor, 2023.

Legenda. Interações (3D e 2D), as pontes de hidrogênio estão destacadas em verde, as interações hidrofóbicas são destacadas em rosa e as interações estéricas são destacadas em vermelho. ALA: Alanina; GLU: Ácido Glutâmico; ILE Isoleucina; LEU Leucina; MET Metionina; PHE Fenilalanina; TRP Triptofano; TYR Tirosina; VAL Valina; SER: Serina; CYS: cisteína; ILE: Isoleucina; ASP: Ác. Aspártico; HIS: Histidina.

De maneira geral, observou-se que a maioria das interações entre o ligante das proteínas selecionadas e as moléculas de menor energia estava relacionada principalmente com ligações de hidrogênio e interações hidrofóbicas. Na enzima cruzaina, notou-se que ela foi a única

macromolécula na qual o ligante e a molécula de maior score apresentaram os cinco tipos de interações estudados, abrangendo ligações de hidrogênio, interações eletrostáticas, ligações covalentes, interações hidrofóbicas e interações estéricas.

Interações entre a enzima tripanotona redutase (PDB: 1GXF) e o composto 149 (Magnoflorina) consistem principalmente em pontes de hidrogênio envolvendo os resíduos Glu19 e Ser15, consideradas interações polares. Essas interações são altamente relevantes, uma vez que as pontes de hidrogênio representam ligações fortes e muito expressivas. Além disso, duas interações hidrofóbicas foram estabelecidas: uma com os resíduos Trp22 e outra com os resíduos Met114. Esses aminoácidos são caracterizados como ligações apolares. Ao comparar a interação do composto 149 com a do ligante PDB [Quinacrine mustard], foi notada coincidência entre os resíduos de ligação, através dos resíduos Glu19, Met114 e Trp22.

Para a enzima Pteridine reductase 2 (PDB: 1MXF), a interação com o composto 038 (Bisdehydroprotostemonine) estabeleceu ligações de hidrogênio envolvendo os átomos de oxigênio (O), consiste principalmente em pontes de hidrogênio envolvendo os resíduos Ser103 e Arg22. Uma interação estérica foi observada, com Tyr105, foram encontradas com os elementos da cadeia carbônica alifática, compreendendo a interação do tipo ligação de hidrogênio. Quando confrontado com a interação do composto 038 com o do ligante PDB [Methotrexate], notou-se uma coincidência entre os resíduos de ligação envolvendo tanto Tyr105 quanto Ser103.

A proteína Dihydrofolate reductase-thymidylate synthase - DHFR (PDB: 3HBB) interagiu com o composto 149 (Magnoflorine), através de cinco interações hidrofóbicas com Pro85, Ile84, Met49, Phe52, Ala28 e Asp48, principalmente através dos anéis aromáticos e ligações de hidrogênio com Ile154 e Val26, e também através dos átomos de hidrogênio (H) do grupo amina (NH). Ao comparar a interação do composto 149 com o ligante PDB [Trimetrexate], foi notada coincidência entre os resíduos de ligação, incluindo os resíduos Ile154, Val26, Asp48, Ala28, Ile84.

A proteína Cruzaina (PDB: 3LXS) interagiu com o composto 367 (N-metil-tiramina) através de uma interação estérica com Gly23, ligação estabelecida através do átomo de oxigênio, assim como ocorreram interações hidrofóbicas com os resíduos de aminoácidos Ser64, Cys25 e His162, nas quais ocorreram principalmente com os anéis aromáticos, e uma delas com átomo de oxigênio do grupo éter metílico (Ser64). Em relação às ligações de hidrogênio, as mesmas foram estabelecidas com os resíduos Cys63 e Asp161. Ao analisar de maneira comparativa a interação do composto 367 com o ligante do PDB [Inibidor piperazina],

notou-se uma coincidência entre os resíduos de ligação, através dos resíduos His162, Cys25 e Asp161.

De maneira geral, algumas enzimas podem não apresentar todos os resíduos que coincidem com a molécula em estudo. No entanto, essa ocorrência é possível e pode indicar diferenças no local de ligação ao sítio ativo da enzima entre o ligante e o composto em análise, mesmo compartilhando o mesmo sítio ativo. As interações podem ocorrer de forma variada, dependendo da conformação da molécula e dos elementos que se ligam aos resíduos de aminoácidos, entre outros fatores.

Foi conduzida uma análise de probabilidade combinada, visando identificar potenciais alvos múltiplos compostos. Esses alvos foram selecionados com base nos resultados do modelo de RF e do *docking* molecular, revelando potenciais atividades prováveis para mais de uma enzima. A fórmula utilizada foi a seguinte:

$$\text{ProbComb} = \frac{(\text{ProbDc} + (1 + \text{ESP}) \times \text{PActivity})}{2 + \text{ESP}}, \text{Se } \text{ProbComb} > 0,5$$

Nesta fórmula, ProbDc representa a probabilidade potencial ativada obtida pela análise de acoplamento molecular, ESP indica o valor médio específico do modelo de RF (valor de especificidade), e PActivity é o valor de probabilidade potencial ativada do modelo de RF. A probabilidade combinada foi ajustada para considerar apenas moléculas com valores acima de 0,5, sendo estas as consideradas prováveis de serem ativas. Os valores de probabilidade combinados foram calculados para os compostos identificados em relação a cada enzima-alvo, e foram analisadas quais moléculas eram *multitarget*.

Foram calculados os valores de probabilidade combinada para as 21 moléculas que apresentaram atividade com probabilidade acima de 70% no modelo de predição; dessa forma, foram executadas em cada enzima investigada. Em seguida, procedeu-se à análise das moléculas *multitarget*, identificando quais se encaixavam nessa categoria ao final do estudo. A probabilidade combinada está sujeita a condições, de forma que apenas as moléculas com valores acima de 0,5 serão consideradas potencialmente ativas. Os cálculos das probabilidades foram realizados para as 21 moléculas em cada enzima estudada, sendo analisado, em seguida, quais delas possuíam a capacidade de atingir múltiplos alvos simultaneamente.

Dos 21 compostos moleculares analisados, todos apresentaram probabilidade de atividade potencial para todas as enzimas estudadas, sugerindo que esses compostos possuem um potencial *multitarget*. O conceito de potencial *multitarget* refere-se a compostos projetados

para modular vários alvos relevantes para uma doença, com o objetivo de aumentar a eficácia ou melhorar a segurança [263].

Os valores da probabilidade combinada podem ser consultados nas tabelas (9), (10), (11) e (12).

Tabela 9. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 1GXF.

ID COMPOSTO	(p) Modelo	(p) Docking	(p) Combinado
149	0.888	0.676426763	0.783477252
194	0.864	0.465923829	0.696141938
150	0.86	0.652168813	0.751184132
333	0.832	0.571905929	0.70243497
320	0.808	0.551804832	0.676264207
326	0.796	0.578214091	0.67386774
114	0.784	0.522141415	0.647717304
222	0.772	0.424914327	0.61049904
319	0.768	0.51864558	0.633521937
367	0.76	0.567240487	0.640357501
38	0.756	0.649496664	0.659608395
115	0.744	0.526482803	0.615845473
151	0.744	0.495556844	0.607475552
152	0.744	0.486593073	0.605049562
309	0.744	0.808615494	0.692202957
315	0.736	0.550899987	0.615760782
369	0.736	0.551526646	0.61592856
153	0.732	0.550064334	0.61219174
350	0.732	0.550415436	0.612285231
2	0.704	0.628321768	0.60881576
329	0.704	0.584768882	0.597662166

Fonte: Próprio autor, 2023.

Tabela 10. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 1MXF.

ID COMPOSTO	(p) Modelo	(p) Docking	(p) Combinado
149	0.888	0.87357355	0.847160899
194	0.864	0.64858515	0.75355168
150	0.86	0.887974181	0.824953727
333	0.832	0.950060595	0.816885564
320	0.808	0.735091178	0.730136658
326	0.796	0.792016129	0.735776224
114	0.784	0.597338396	0.669163078
222	0.772	0.594334506	0.658077206

319	0.768	0.729448195	0.692414774
367	0.76	0.829985248	0.712997013
38	0.756	0.933459198	0.737700674
115	0.744	0.694985575	0.661449715
151	0.744	0.896688546	0.716039388
152	0.744	0.740238893	0.673697249
309	0.744	0.722446031	0.66888172
315	0.736	0.837178679	0.692407242
369	0.736	0.837095195	0.69238489
153	0.732	0.735701211	0.661622876
350	0.732	0.694786083	0.650728051
2	0.704	0.71563978	0.631177303
329	0.704	0.806962026	0.654564302

Fonte: Próprio autor, 2023.

Tabela 11. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 3HBB.

ID COMPOSTO	(p) Modelo	(p) Docking	(p) Combinado
149	0.888	0.992547744	0.885592724
194	0.864	0.752727351	0.786283169
150	0.86	0.865133889	0.817808347
333	0.832	0.602051932	0.711558824
320	0.808	0.82314322	0.75601736
326	0.796	0.813850401	0.742098552
114	0.784	0.586590903	0.666097951
222	0.772	0.642677998	0.671653479
319	0.768	0.836910054	0.722436857
367	0.76	0.719177905	0.682362753
38	0.756	0.317399035	0.568278527
115	0.744	0.654334186	0.650447666
151	0.744	0.870048459	0.708829412
152	0.744	0.78933348	0.686984398
309	0.744	0.859758511	0.706044501
315	0.736	0.854189824	0.696961699
369	0.736	0.854410507	0.697020783
153	0.732	0.762860505	0.668854816
350	0.732	0.814646277	0.682644261
2	0.704	0.944713251	0.689841444
329	0.704	0.836345846	0.662089296

Fonte: Próprio autor, 2023.

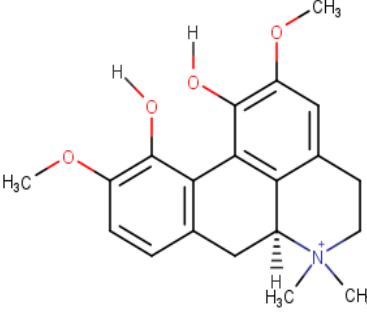
Tabela 12. Valores obtidos na probabilidade combinada de Docking Molecular e Modelo de predição para proteína 3LXS.

ID COMPOSTO	(p) Modelo	(p) Docking	(p) Combinado
149	0.888	0.75392011	0.808509662
194	0.864	0.62969903	0.747615848
150	0.86	0.74140498	0.779100868
333	0.832	0.92813999	0.810251173
320	0.808	0.82677019	0.757083417
326	0.796	0.77727306	0.731507223
114	0.784	0.86549146	0.745638887
222	0.772	0.72899431	0.695893636
319	0.768	0.83737665	0.722567213
367	0.76	0.94741716	0.745462729
38	0.756	0.61421341	0.649905178
115	0.744	0.80798343	0.692031892
151	0.744	0.80973788	0.692506724
152	0.744	0.86062433	0.706278828
309	0.744	0.8562381	0.705091723
315	0.736	0.895938	0.708139095
369	0.736	0.80386517	0.683488092
153	0.732	0.79276037	0.676816511
350	0.732	0.55827542	0.614378178
2	0.704	0.9069106	0.680160445
329	0.704	0.86197422	0.668652548

Fonte: Próprio autor, 2023.

Visando filtrar as moléculas com maior potencial *multitarget*, foi selecionada a molécula que apresentou o maior percentual em todas as enzimas estudadas. Dessa forma, foi escolhida a molécula 149, pois apresentou valores percentuais de 78% para a enzima 1GXF. Em relação à enzima 1MXF, essa molécula demonstrou percentuais de 84%. Já para a enzima 3HBB, os valores percentuais foram de 88%, e finalmente, para a enzima 3LXS, alcançou percentuais de atividade de 80%. A estrutura molecular desse composto em 2D, juntamente com as enzimas em que ele pode exercer atividade, pode ser visualizada na tabela 13.

Tabela 13. Molécula com Potencial Ativo e *multitarget*.

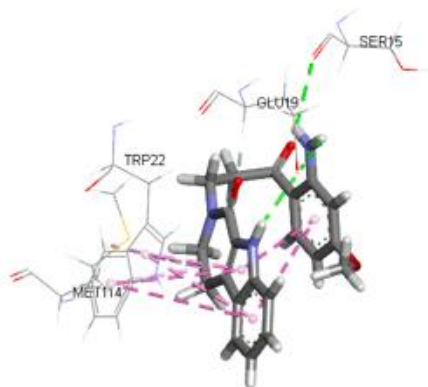
Molécula potencialmente ativa	Gênero do Alcaloide	Arquitetura molecular em 2D	Proteínas
Magnoflorine	Croton		1GXF 1MXF 3HBB 3LXS

Fonte: Próprio autor, 2023.

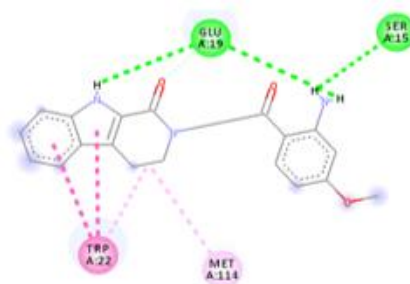
A figura 12 exibe a representação da molécula Magnoflorina ligando-se ao sítio ativo das quatro proteínas, evidenciando as interações presentes no complexo proteína-ligante. A (S)-magnoflorina é um alcaloide aporfínico, um tipo de anel aromático heterocíclico, no qual o nitrogênio foi quaternizado por um grupo metil adicional. Possui um íon amônio quaternário. Resumidamente, a estrutura básica consiste em quatro anéis, sendo três anéis benzênicos fundidos a um anel de sete membros contendo um átomo de nitrogênio. Além disso, possui grupos funcionais, como hidroxilas (-OH), que podem atuar como doadores ou aceitadores de ligações de hidrogênio. Esses grupos podem interagir com resíduos aminoácidos na enzima, que possuem átomos de hidrogênio disponíveis para formar essas ligações, possibilitando interações com as diversas enzimas estudadas. A sua nomenclatura "(S)" indica a configuração espacial do composto, relacionada com a disposição dos grupos químicos ao redor do átomo de carbono assimétrico presente na molécula. Essas características podem explicar sua habilidade de se conectar a todas essas proteínas.

Enzimas: PDB: 1GXF

A



B

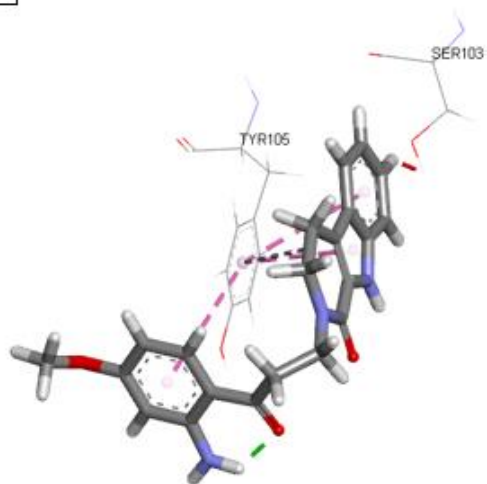


Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Pi Stacked
- Alkyl
- Pi-Alkyl

Enzimas: PDB: 1MXF

A



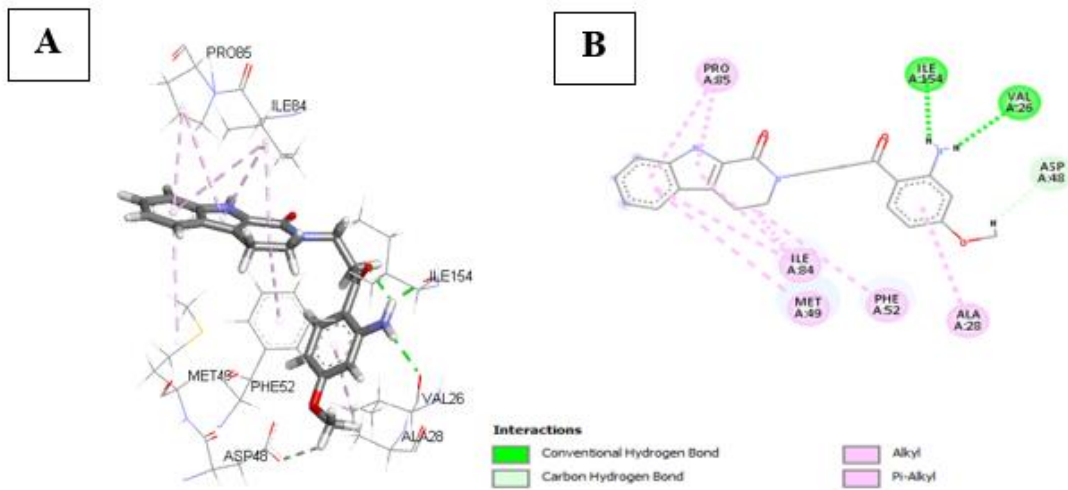
B



Interactions

- Unfavorable Bump
- Pi-Pi Stacked
- Pi-Pi T-shaped
- Pi-Alkyl

Enzimas: PDB: 3HBB



Enzimas: PDB: 3LXS

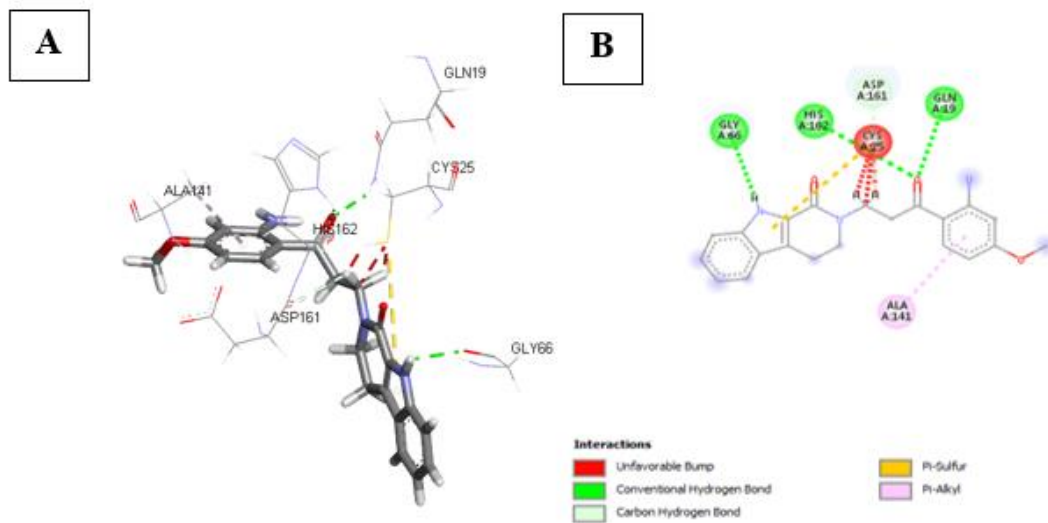


Figura 12. Interações realizadas pelo composto Magnoflorine com os resíduos de aminoácidos correspondentes, proteínas 1GXF; 1MXF; 3HBB; 3LXS, em 3D (A) e 2D (B).

Fonte: Próprio autor, 2023.

Legenda. Interações (3D e 2D), as pontes de hidrogênio estão destacadas em verde, as interações hidrofóbicas são destacadas em rosa e as interações estéricas são destacadas em vermelho. ALA: Alanina; GLU: Ácido Glutâmico; ILE Isoleucina; LEU Leucina; MET Metionina; PHE Fenilalanina; TRP Triptofano; TYR Tirosina; VAL Valina; SER: Serina; CYS: cisteína; ILE: Isoleucina; ASP: Ác. Aspártico; HIS: Histidina.

Ao observar as interações realizadas pelo composto Magnoflorine, percebe-se que os átomos que interagem com resíduos de aminoácidos correspondem, em sua maioria, aos

elementos hidrogênio, carbono dos anéis benzênicos e oxigênio. As interações ocorridas abrangem os resíduos de interações hidrofóbicas Trp22 e Met114, bem como ligações de hidrogênio Glu19 e Ser15, para a enzima de PDB:1GXF. Para a enzima 1MXF, ocorrem nas interações hidrofóbicas nos resíduos Tyr105 e nas ligações de hidrogênio Ser103. Na enzima 3HBB, as interações abrangem ligações de hidrogênio com resíduos Ile110 e Val26, bem como interações hidrofóbicas em Pro85, Ile84, Met49, Phe52, Ala28 e Asp48. Para as enzimas 3LXS, ocorrem em resíduos de ligação de hidrogênio Gly66, His162, Cys25, Gln19 e interações hidrofóbicas em Ala141.

6.4 Previsão de Propriedades ADMET

O conjunto de 21 alcaloides que apresentaram resultados promissores, de acordo com probabilidade combinada de docking molecular e modelo de predição (potencial *multitarget*) com percentuais maiores que 59% foram submetido a uma avaliação de diversos parâmetros preditivos para selecionar os compostos com os melhores perfis ADMET. Com base nas propriedades físico-químicas, foi realizada uma análise da absorção de cada composto, utilizando a regra de Lipinski como parâmetro. Para isso, as moléculas precisaram atender aos seguintes critérios: o número máximo de doadores de ligação de hidrogênio no composto não deve exceder cinco, o número máximo de aceptadores de ligação de hidrogênio não deve ser superior a dez, o valor do logP deve ser menor ou igual a cinco e o peso molecular não deve ultrapassar 500 g/mol [131]. Dessa forma, os compostos que não violam esses parâmetros possuem excelentes características de absorção e biodisponibilidade.

Considerando que os fármacos atualmente utilizados para o tratamento da doença de Chagas são administrados principalmente por via oral, por ser a rota mais comum e conveniente para a administração de medicamentos, já que muitos fármacos para o tratamento da doença de Chagas estão disponíveis em forma de comprimidos ou cápsulas. Dessa forma, torna-se necessário analisar a biodisponibilidade oral e a taxa de absorção das moléculas candidatas. Neste estudo, observou-se que 20 compostos atenderam aos parâmetros da regra de Lipinski, não apresentando nenhuma violação. No entanto, o composto de ID 367 apresentou duas violações a esse critério.

A capacidade de dissolução em lipídios (lipofilicidade) e em água (solubilidade), pode influenciar a distribuição de um medicamento dentro do organismo [261]. O coeficiente de partição mais comumente utilizado para avaliar a lipofilicidade é o log P, que representa a razão da concentração de um composto em n-octanol e em água. Valores de logP acima de 5.0

geralmente indicam alta lipofilicidade. Observou-se que todos os 21 compostos apresentaram boa lipofilicidade, sendo que grande parte foi considerada moderadamente solúvel.

É de extrema importância ressaltar que os candidatos devem apresentar valores de logP equilibrados, evitando tanto valores muito altos quanto muito baixos. Isso se deve ao fato de que substâncias altamente hidrofílicas tendem a permanecer em solução aquosa, dissolvendo-se rapidamente, mas não conseguem atravessar as membranas biológicas, que são ricas em lipídios. Por outro lado, substâncias altamente hidrofóbicas dissolvem-se nas membranas e nelas se acumulam, resultando em efeitos cumulativos e gerando toxicidade. É essencial encontrar um equilíbrio para garantir a adequada absorção e distribuição dos candidatos no organismo, evitando problemas de eficácia e segurança [258].

A permeabilidade dos candidatos através da membrana biológica é influenciada pelo peso molecular e pela área de superfície polar topológica (TPSA). Se um candidato tiver um peso molecular e TPSA altos, isso pode resultar em uma redução da sua permeabilidade pela membrana. Isso ocorre porque se pressupõe que uma molécula grande e pesada terá dificuldade em passar pela membrana. Por outro lado, candidatos com peso molecular e TPSA baixos têm uma maior probabilidade de serem permeáveis nas membranas biológicas [259]. Além disso, de acordo com a literatura, candidatos com áreas de superfície polar topológica (TPSA) menores ou iguais a 140 Å (ångströms) têm uma melhor biodisponibilidade oral [260]. Nesse estudo, 20 compostos, com exceção do composto de ID 367, apresentaram TPSA menor que 140 Å. Isso significa que esses compostos podem ser considerados candidatos com uma maior capacidade de serem absorvidos pelo organismo quando administrados por via oral.

O processo metabólico (metabolismo) pode ter impactos significativos na eficácia de um medicamento, influenciando a biodisponibilidade, a meia-vida e a formação de metabólitos prejudiciais ou sem potencial terapêutico. A análise dos resultados demonstrou que os compostos com ID 150, 333, 115 e 151 apresentaram os melhores perfis farmacocinéticos.

A tabela 14 apresenta o perfil ADMET de 21 alcaloides selecionados com base na probabilidade combinada de docking molecular e no modelo de predição (potencial *multitarget*).

Tabela 14. Propriedades físico-químicas e farmacocinéticas do ADMET para os alcaloides.

Moléculas (ID)	LogP Consensus	TPSA	ABS %	Violação a regra de Lipinski	GI*	BBB*	P-gp*	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4
149	0.65	58.92	88.33	0	Alta	Sim	Sim	Sim	Não	Não	Não	Sim
194	1.64	25.78	100.10	0	Alta	Sim	Não	Não	Não	Não	Não	Não
150	2.68	88.42	78.49	0	Alta	Não	Sim	Sim	Sim	Sim	Sim	Sim
333	2.82	66.59	86.02	0	Alta	Sim	Sim	Sim	Sim	Sim	Sim	Sim
320	3.37	91.00	77.60	0	Alta	Não	Não	Sim	Sim	Sim	Sim	Não
326	2.38	37.91	95.92	0	Alta	Sim	Não	Sim	Não	Não	Sim	Sim
114	2.21	60.16	88.24	0	Alta	Sim	Não	Não	Não	Não	Não	Não
222	1.75	60.39	88.16	0	Alta	Sim	Não	Não	Não	Não	Sim	Não
319	3.37	91.00	77.60	0	Alta	Não	Não	Sim	Sim	Sim	Sim	Não
367	3.79	154.10	55.83	2	Baixo	Não	Sim	Não	Não	Não	Não	Sim
38	2.67	74.30	83.36	0	Alto	Sim	Não	Não	Não	Não	Não	Não
115	3.70	46.17	93.07	0	Alto	Sim	Não	Sim	Sim	Sim	Sim	Não
151	2.69	88.42	78.49	0	Alto	Não	Sim	Sim	Sim	Sim	Sim	Sim
152	3.88	50.80	91.47	0	Alto	Sim	Sim	Sim	Sim	Não	Sim	Sim
309	2.58	94.58	76.36	0	Alto	Não	Sim	Sim	Não	Não	Sim	Não
315	3.44	44.05	93.80	0	Alto	Sim	Sim	Sim	Sim	Não	Sim	Sim
369	2.30	48.00	92.44	0	Alta	Sim	Não	Não	Não	Sim	Não	Sim
153	3.87	50.80	91.47	0	Alta	Sim	Sim	Sim	Sim	Não	Sim	Sim
350	2.57	85.35	79.55	0	Alta	Não	Não	Sim	Não	Não	Sim	Não
2	2.64	65.07	86.55	0	Alta	Sim	Sim	Não	Não	Não	Não	Não
329	3.47	58.22	88.91	0	Alta	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Legendas: LogP: coeficiente de partição; TPSA: área da superfície polar topológica; ABS %: percentagem de absorção oral; GI: absorção gastrointestinal; BBB: barreira hematoencefálica; P-glicoproteína; CYP: Modelos de metabolismo de citocromo; P.M: Peso Moléculas.

*Os nomes dos compostos estão no anexo I.

Fonte: Próprio autor, 2023.

Durante a fase de pesquisa e desenvolvimento de novos fármacos, muitos candidatos são eliminados devido a problemas associados, principalmente à toxicidade. A toxicidade desempenha um papel crucial no processo de desenvolvimento e planejamento de candidatos a fármacos, sendo um dos parâmetros essenciais da análise ADMET [241]; [242]. Juntamente com a eficácia da ação terapêutica, a toxicidade permite estimar as doses e investigar a viabilidade de utilização do candidato como fármaco, além de explorar possíveis combinações de acordo com as necessidades específicas [243]; [244]. Para análise dos riscos de toxicidade, foram selecionados os compostos que apresentaram probabilidade de atividade igual ou superior a 59% na probabilidade combinada. Nesse contexto, com base nas análises conduzidas utilizando o OSIRIS Data Warrior, foram identificados três métodos de predição de riscos, sendo eles a mutagenicidade, irritabilidade do tecido e toxicidade no sistema reprodutor.

A Mutagenicidade é a capacidade de uma substância causar mutações no material genético de um organismo vivo. As mutações são alterações na sequência de DNA que podem ocorrer espontaneamente ou ser induzidas por agentes mutagênicos, como certos produtos químicos, radiação ou agentes físicos [255]. No que diz respeito à toxicidade no sistema reprodutivo, esse método permite avaliar a capacidade da substância em causar interferências no sistema reprodutivo humano [256]. Em relação à análise da irritabilidade do tecido, a mesma corresponde à capacidade de uma substância causar irritação na pele ou em outros tecidos do corpo humano. A irritação é uma resposta inflamatória localizada que pode ocorrer quando uma substância entra em contato com a pele, membranas mucosas ou outros tecidos [257]. Dessa forma, na tabela 15 estão os dados da análise de toxicidade das 211 moléculas que demonstraram probabilidades de atividade acima de 59%.

Tabela 15. Resultado das análises de toxicidade dos alcaloides em estudo.

Moléculas (ID)	MUT	TUM	ESR	IRR	TOX
149	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
194	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
150	Baixo	Alto	Nenhum	Nenhum	Sim
333	Nenhum	Nenhum	Alto	Nenhum	Sim
320	Baixo	Nenhum	Nenhum	Nenhum	Sim
326	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
114	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
222	Alto	Nenhum	Nenhum	Nenhum	Sim
319	Alto	Nenhum	Nenhum	Nenhum	Sim
367	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
38	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
115	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
151	Nenhum	Nenhum	Alto	Nenhum	Sim
152	Baixo	Nenhum	Nenhum	Nenhum	Sim
309	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
315	Nenhum	Nenhum	Alto	Nenhum	Sim
369	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
153	Baixa	Nenhum	Nenhum	Nenhum	Sim
350	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
2	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
329	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum

Legendas: MUT: mutagenicidade; TUM: tumorigenicidade; ESR: efeito no sistema reprodutor; IRR: irritabilidade no tecido; TOX: toxicidade geral frente a esses 4 parâmetros.

Fonte: Próprio autor, 2023.

Baseado nos dados da tabela acima, as moléculas de ID 150, 333, 320, 222, 319, 151, 152, 315 e 153 demonstraram riscos em pelo menos um dos parâmetros de toxicidade. De acordo com a literatura, alguns alcaloides são conhecidos por causar efeitos tóxicos em humanos e animais, enquanto muitos outros têm sido utilizados na medicina devido às suas propriedades terapêuticas [178; 165]. Essa justificativa apoia sua contínua utilização e aplicação em pesquisas para a descoberta de novos agentes terapêuticos, incluindo analgésicos, antitumorais, entre outros [121]. Sendo assim, os compostos moleculares que apresentaram toxicidade em pelo menos 1 (um) dos parâmetros, dessa forma podem ser considerados para estudos futuros.

6.5 Dinâmica Molecular

A métrica do desvio quadrado médio (RMSD) foi calculada para os átomos de Ca em cada enzima complexada e as estruturas de cada ligante, separadamente. Análise de desvio quadrado médio (RMSD) do complexo de proteínas Diidrofolato redutase-timidilato sintase de *Trypanosoma cruzi* (PDB: 3HBB) [265] com o composto teste Magnoflorine (linha vermelha)

mostrou que a enzima atingiu conformações variando de 4,0 a 6,7 nm de tamanho em 100 ns (Figura 13). Foi observada uma estabilidade no complexo formado pelo alvo em estudo com o composto Magnoflorine como também nos complexos da macromolécula e do ligante co-cristalizado, porém o composto teste apresentou valores de RMSD mais altos quando comparado a macromolécula Diidrofolato redutase-timidilato sintase de *Trypanosoma cruzi* (PDB: 3HBB) [265] (linha preta) e ao complexo da enzima em estudo com o ligante co-cristalizado (linha verde).

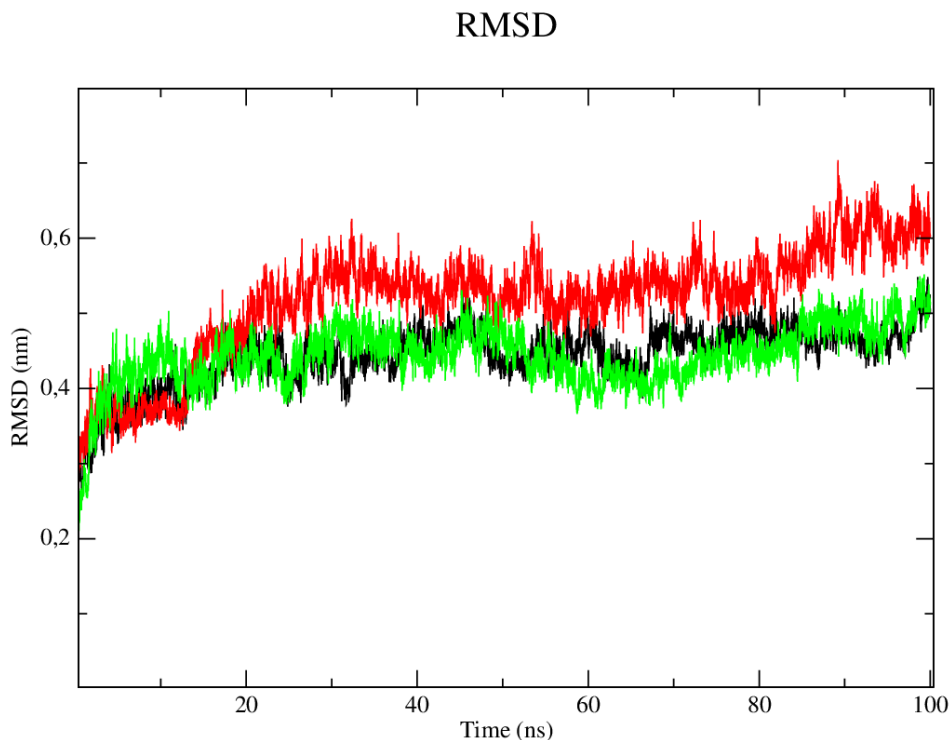


Figura 13. RMSD de átomos C α da proteína Diidrofolato redutase-timidilato sintase de *Trypanosoma cruzi* (PDB: 3HBB) (linha preta) e complexado ao composto de teste Magnoflorine (linha vermelha) e ao ligante PDB Trimetrexato (linha verde).

Fonte: Próprio autor, 2023.

Quando se analisa a flexibilidade dos ligantes (Figura 14), observa-se que o composto teste Magnoflorine (linha preta) apresentou maior estabilidade, pois este apresentou menores flutuações nos valores de RMSD no período de 100 ns da simulação quando comparado com o ligante co-cristalizado Trimetrexato (linha vermelha). O ligante PDB Trimetrexato foi drasticamente mais instável, este apresentou picos de instabilidade entre 20 e 50 ns, sendo observada a ocorrência de valores de RMSD correspondentes a 3.3 nm.

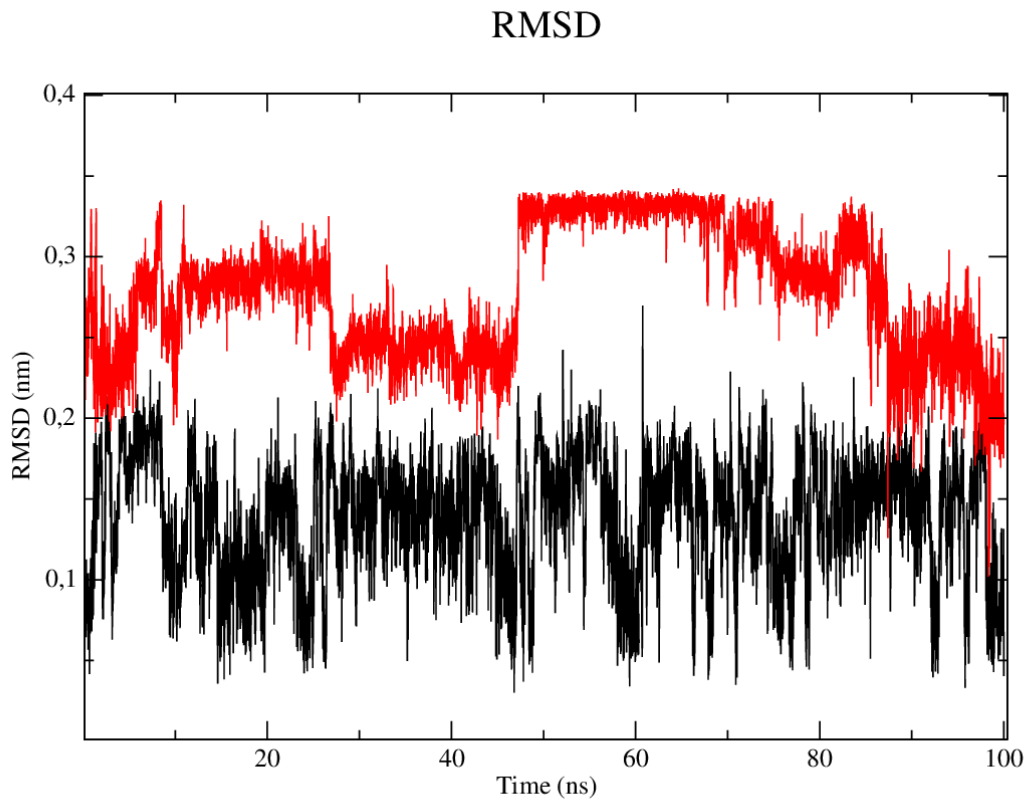


Figura 14. RMSD dos átomos $C\alpha$ dos compostos. Composto de teste Magnoflorine (linha preta) e ligante PDB Trimetrexato (linha vermelha).

Fonte: Próprio autor, 2023.

As flutuações da raiz quadrada média individual (RMSF) também foram calculadas como resíduos (Figura 15). Os valores de RMSF refletem a diminuição da flexibilidade. Dado que os aminoácidos com flutuações acima de 0,3 nm contribuem para a flexibilidade da estrutura da proteína, verificou-se que no sítio alostérico os resíduos nas posições: 2, 515 e 516 contribuem para a flexibilidade da proteína.

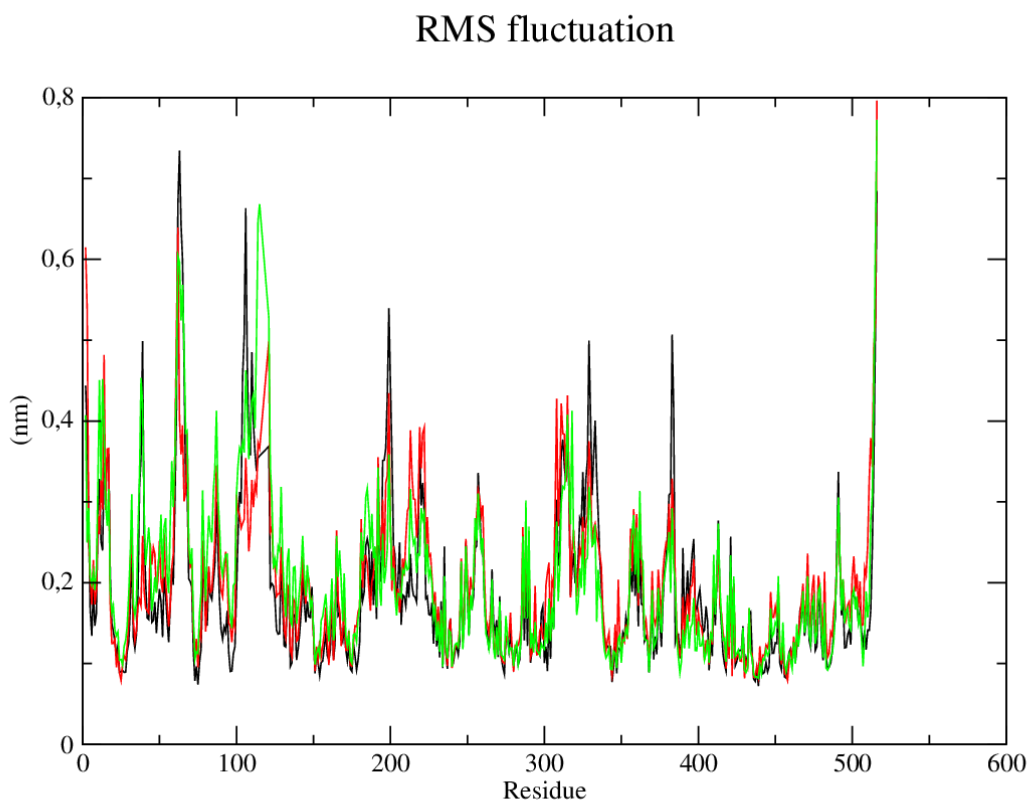


Figura 15. RMSF de átomos. Alvo Diidrofolato redutase-timidilato sintase de *Trypanosoma cruzi* (PDB: 3HBB) (linha preta) complexado ao composto de teste Magnoflorine (linha vermelha) e ligante PDB Trimetrexato (linha verde).

Fonte: Próprio autor, 2023.

As energias de interação de Coulomb e Lennard-Jones (tabela 16) dos complexos proteína-ligante foram calculadas e o composto de teste magnoflorine demonstrou ter uma menor estabilidade de interação com o ligante PDB Trimetrexato de acordo com as métricas de Coulombic e Lenard-Jonnes.

Tabela 16. Valores de energia de interação de Coulombic e Lennard-Jones.

Ligantes	Diidrofolato redutase-timidilato sintase de <i>Trypanosoma cruzi</i> (PDB: 3HBB)	
	Coul	LJ
Magnoflorine	-67.3845	-138.461
Ligante PDB	-103.806	-171.116

Fonte: Próprio autor, 2023.

As análises demonstram que o RMSD do complexo para o composto teste Magnoflorina apresentou estabilidade, além disso, pode-se sugerir que através de cálculos de RMSF e energia de interação, o composto teste Magnoflorine interage com este alvo, pois permite interação, flexibilidade e estabilidade.

7 CONCLUSÕES

A partir das informações obtidas pela triagem virtual, os compostos avaliados demonstraram um potencial promissor contra o *T. cruzi*, o que justifica a realização do teste biológico posteriormente para validar esses resultados obtidos por meio da simulação *in silico*. Neste estudo, foram analisadas as interações moleculares de alcaloides pertencentes à família Euphorbiaceae com diferentes proteínas-alvo potenciais para o desenvolvimento de drogas contra o *T. cruzi*, sendo que esses estudos foram conduzidos por meio de modelos de predição, docagem molecular, simulação de dinâmica molecular e estudos de energia de ligação.

Através das técnicas *in silico* aplicadas neste estudo, conseguimos desenvolver modelos para realizar um rastreamento virtual do banco de dados de alcaloides pertencentes à família Euphorbiaceae. Observou-se que o modelo preditivo demonstrou confiabilidade na identificação das moléculas ativas e inativas, possibilitando, assim, a progressão do processo de triagem virtual. O modelo elaborado para a forma tripomastigota identificou cerca de 91 moléculas com potencial atividade, apresentando probabilidade acima de 50%. Além disso, utilizando descritores VolSurf, o modelo aplicado à forma tripomastigota do *T. cruzi* selecionou 21 moléculas com potencial atividade, com uma probabilidade superior a 70%.

A avaliação da docagem molecular, utilizando a abordagem baseada na estrutura do ligante no software Molegro Virtual Docking, foi conduzida com quatro enzimas específicas para o *T. cruzi*. Essa avaliação determinou o potencial antiprotozoário de 21 moléculas que alcançaram um percentual de atividade de 70% no modelo de predição. As energias de ligação foram calculadas através das funções de pontuação Moldock, PLANTS e Rerank, levando à identificação de 18 compostos com potencial de atividade contra a enzima 1GXF. Além disso, para as enzimas 1MXF, 3HBB e 3LXS, todas as 21 moléculas demonstraram potencial de atividade.

Ao combinar ambas as abordagens (resultados do docking molecular com o modelo RF) para *T. cruzi*, todas as 21 moléculas analisadas mostraram probabilidade de atividade em relação a todas as enzimas estudadas, sugerindo um potencial *multitarget* desses compostos. Para aprimorar a seleção das moléculas com maior potencial *multitarget*, a molécula que

apresentou o maior percentual de atividade em todas as enzimas estudadas foi escolhida; assim, a molécula 149 (Magnoflorina) foi selecionada.

Em geral, os compostos demonstraram previsões de alta absorção via oral. Os resultados *in silico* indicaram taxas de absorção superiores a 77% para a maioria dos compostos, apresentando, na maioria dos casos, apenas uma violação da regra de Lipinski. Isso sugere uma boa disponibilidade. Além disso, somente o composto com o ID 367 apresentou indicações de toxicidade em dois dos parâmetros avaliados. Consequentemente, os alcaloides analisados exibiram perfis satisfatórios de biodisponibilidade, de acordo com os critérios da regra de Lipinski, e a maioria deles não demonstrou riscos de toxicidade nos parâmetros avaliados, sugerindo, assim, um perfil positivo de ADMET para a série em estudo.

A estabilidade da interação foi avaliada por meio de simulações de dinâmica molecular, evidenciando que a Magnoflorina manteve-se estável. As flutuações observadas no RMSF indicam que o composto de teste, a Magnoflorina, interage com o alvo, demonstrando sua capacidade de interação, flexibilidade e estabilidade.

Sugerimos, portanto, como perspectiva futura, encaminhar a Magnoflorina para validação experimental do método *in silico* por meio de testes biológicos. Os resultados obtidos na triagem virtual demonstram que este composto se apresenta como um candidato potencial com atividade anti- *T. cruzi*.

REFERENCIAS

- [1] CONCEIÇÃO, J. R. *et al.* Neglected tropical diseases and systemic racism especially in Brazil: from socio-economic aspects to the development of new drugs. **Acta Tropica**, 2022. p. 106654.
- [2] RODRIGUES, M. D.; GOMIDE, M. Acesso à fitoterapia na atenção básica em saúde através da Análise de Redes Sociais (ARS). **Redes: Revista Hispana para el análisis de redes sociales**, 2019. 30, 244-253.
- [3] FELIX, E. M. *et al.* Mercado de Produtos de Controle de peso e fitoterápicos: uma análise comparativa entre o Brasil e o Mercosul. **Revista Interdisciplinar Encontro das Ciências RIEC**, 2019. [S.L.], v. 2, n. 1, p. 489-495.
- [4] NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. **Journal Of Natural Products**, 2020. [S.L.], v. 83, p. 770-803.
- [5] AMMAR, O. *In silico* pharmacodynamics, toxicity profile and biological activities of the Saharan medicinal plant *Limoniastrum feei*. **Brazilian Journal of Pharmaceutical Sciences**, 2017. [S.L.], v. 53, n. 3, p. 1-10.
- [6] SILVA, J. L. *et al.* Introdução a Triagem Virtual. **Bioinfo - Revista Brasileira de Bioinformática e Biologia Computacional**, 2021. [S.L.], v. 1, p. 158-174.
- [7] OLIVEIRA, J. R. G. **Estudo fitoquímico e avaliação do efeito citotóxico de *Cnidoscolus quercifolius* Pohl (Euphorbiaceae) em células do melanoma humano**. Dissertação (mestrado em recursos naturais) Universidade Federal do Vale do São Francisco, Petrolina, 2017.
- [8] OLIVEIRA, J. R. G. *et al.* Antibacterial activity of terpenoids isolated from *Cnidoscolus quercifolius* Pohl (Euphorbiaceae), a Brazilian medicinal plant from Caatinga biome. **European Journal of Integrative Medicine**, 2018. 24, pp.30-34.
- [9] TRINDADE, M. J. S.; LAMEIRA, O.A. Species from the Euphorbiaceae family used for medicinal purposes in Brazil. *Rev. Cubana Plan. Medicinales*, 2014. v.19, p.292–309.
- [10] FERNANDES, A. F. C. **Physical-Chemical Characterization of Plant Drug, Phytochemical and Pharmacological Study of *Cnidoscolus Quercifolius* Pohl (Euphorbiaceae)**. 139 f. Dissertação (Mestrado) - Center for Biological and Health Sciences, State University of Paraíba, Campina Grande, 2019.
- [11] ATANASOV, A. G.; ZOTCHEV, S. B.; DIRSCH, V. M.; SUPURAN, Claudiu T. Natural products in drug discovery: advances and opportunities. **Nature Reviews Drug Discovery**, 2021. [S.L.], v. 20, n. 3, p. 200-216.
- [12] ANAND, U. A. *et al.* Comprehensive Review on Medicinal Plants as Antimicrobial Therapeutics: potential avenues of biocompatible drug discovery. **Metabolites**, 2019. [S.L.], v. 9, n. 11, p. 258.

- [13] MAJOLO, F. et al. Medicinal plants and bioactive natural compounds for cancer treatment: important advances for drug discovery. **Phytochemistry Letters**, 2019. [S.L.], v. 31, p. 196-207.
- [14] MILLER, S. A.; FERREIRA, J.P.; LEJEUNE, J.T. Antimicrobial Use and Resistance in Plant Agriculture: A One Health Perspective. **Agriculture**, 2022, 12, 289.
- [15] JÚNIOR, J. N. S. *et al.* Green microalgae as a potential source of trypanocide compounds. **Natural Product Research**, 2023. p. 1-7.
- [16] MAGOZWI, D.K. *et al.* Flavonóides do gênero euphorbia: Isolamento, estrutura, atividades farmacológicas e relações estrutura-atividade. **Pharmaceuticals**, 2021, 14, 428.
- [17] KEMBOI, D., PETER, X., LANGAT, M., AND TEMBU, J. A review of the ethnomedicinal uses, biological activities, and triterpenoids of Euphorbia species. **Molecules**, 2020. 25:4019
- [18] FAIS, A. *et al.* *Euphorbia characias*: Phytochemistry and Biological Activities. **Plants**. 2021; 10(7):1468.
- [19] SALEHI, B.; IRITI, M.; VITALINI, S.; ANTOLAK, H. Produtos naturais derivados de Euphorbia com potencial para uso na manutenção da saúde. **Biomolecules**, 2019. 9, 337.
- 17 [20] KEMBOI, D.; PEDRO, X.; LANGAT, M.; TEMBU, J. Uma revisão dos usos etnomedicinais, atividades biológicas e triterpenóides de espécies de euphorbia. **Molecules** 2020. 25, 4019.
- [21] RAMALHO, S. *et al.* Biologically Active Orbitides from the Euphorbiaceae Family. **Planta Medica**, 2017. [S.L.], v. 84, n. 09/10, p. 558-567.
- [22] MAGWILU, K. D. *et al.* Phylogeny, Phytomedicines, Phytochemistry, Pharmacological Properties, and Toxicity of *Croton gratissimus* Burch (Euphorbiaceae). **Advances in Pharmacological and Pharmaceutical Sciences**, 2022. [S.L.], v. 2022. p. 1-13.
- [23] PIMENTEL, B. S. *et al.* Taxonomic significance of the distribution of constituents of leaf cuticular waxes of *Croton* species (Euphorbiaceae). **Biochem. Syst. Ecol.** 2020. 92, 104106.
- [24] BEZERRA, F. W. *et al.* Bioactive compounds and biological activity of *Croton* species (Euphorbiaceae): An overview. **Curr. Bioact. Compd.** 2020, 16, 383–393.
- [25] ABREU, L. S. *et al.* Phenylpropanoids from *Croton velutinus* with cytotoxic, trypanocidal and anti-inflammatory activities. *Fitoterapia* 2020, 145, 104632.
- [26] MARTINS, A. O. B. P. B. *et al.* Anti-edematogenic and anti-inflammatory activity of the essential oil from *Croton rhamnifolioides* leaves and its major constituent 1,8-cineole (eucalyptol). **Biomed. Pharmacother.** 2017, 96, 384–395.
- [27] OLIVEIRA, B. P. B. M. A. *et al.* Anti-inflammatory and physicochemical characterization of the *Croton rhamnifolioides* essential oil inclusion complex in β -cyclodextrin. **Biology**, 2020, 9, 114.

- [28] YAO, K. B. *et al.* Phytochemical screening and antibacterial activity of the aqueous extract of the leaves of *Alchornea cordifolia* (Euphorbiaceae) on the in-vitro growth of tetracycline-resistant strains of avian *Escherichia col.* **J Pharmacogn Phytochem**, 2023;12(1):01-05.
- [29] NODJITOLOUM, M. *et al.* Phytochemical screening, phenolic determination and antibacterial activity of the extracts of *Bridelia scleroneura* Muell. Arg. (Euphorbiaceae) from Chad. **J Pharmacogn Phytochem**, 2023;12(2):44-47.
- [30] ATRI, A. *et al.* Green synthesis of copper oxide nanoparticles using *Ephedra Alata* plant extract and a study of their antifungal, antibacterial activity and photocatalytic performance under sunlight. **Heliyon**, 2023. 9(2):e13484.
- [31] ARAÚJO, A.C.J. *et al.* Essential oil of *Croton ceanothifolius* Baill. potentiates the effect of antibiotics against multiresistant bacteria. **Antibiotics**, 2020, 9, 27.
- [32] COSTA, L. S. *et al.* First report on the chemical composition, antioxidant capacity, and preliminary toxicity to *Artemia salina* L. of *Croton campinarenis* Secco, A. Rosário & PE Berry (Euphorbiaceae) essential oil, and *in silico* study. **Antioxidants**. 2022, 11, 2410.
- [33] KAMARUZAMAN, A. S. Chemical Constituents from *Macaranga lowii* King ex Hook.f. (Euphorbiaceae) and Their Effects on Antioxidant Activity. **Asm Science Journal**, 2023. [S.L.], v. 18, p. 1-8.
- [34] OLIVEIRA, S. D. D. S. *et al.* Radical scavenging activity of the essential oils from *Croton grewoides* Baill accessions and the major compounds eugenol, methyl eugenol and methyl chavicol. **J. Essent. Oil Res.** 2021, 33, 94–103.
- [35] RUGAIE, O. A, *et al.* Potencial antimicrobiano, antibiofilme e antioxidante de quatro plantas halófitas, *Euphorbia chamaesyce*, *Bassia arabica*, *Fagonia mollis* e *Haloxylon salicornicum*, crescendo na região de Qassim da Arábia Saudita: perfil fitoquímico e investigações de bioatividade in vitro e *in silico*. **Antibióticos**. 2023; 12(3):501.
- [36] SOUSA, A. *et al.* Antioxidant properties of *Croton zehntneri* Pax et Hoffm. Essential oil and its inclusion complex with β -cyclodextrin prepared by spray drying. **J. Braz. Chem. Soc.** 2022, 33, 1244–1253.
- [37] GUERRA, JUNIOR, J. I. *et al.* *Croton* sp.: A review about popular uses, biological activities and chemical composition. **Res. Soc. Dev.** 2022, 11, e57311225306.
- [38] CHOODEJ, S; PUDHOM, K. Cycloartane triterpenoids from the leaves of *Euphorbia neriifolia*. **Phytochemistry Letters**, 2020. [S.L.], v. 35, p. 1-5.
- [39] BESSIKE, J. G. *et al.* Evaluation of the potentials of *Jatropha curcas* seed oil and in combination with leaf extracts of *Cymbopogon citratus*, *Ocimum basilicum*, and *Eucalyptus globulus* as wood preservatives against *Macrotermes bellicosus* termites. **Industrial Crops and Products**, 2023. [S.L.], v. 195, p. 116205.
- [40] ALYAS, T. *et al.* Applications of scanning electron microscopy in taxonomy with special reference to family Euphorbiaceae. **Microscopy Research and Technique**, 2020. 83, 1066– 1078.

- [41] MOTINIA, N. *et al.* Phytochemical Study of *Euphorbia turcomanica* Boiss. **Metabolites**, 2022. v. 12, n. 12, p. 1200.
- [42] FERNANDES, V. D. G. *et al.* Avaliação *in silico* da atividade antiprotozoário do monoterpeno ascaridol. **Research, Society and Development**, 2020. v. 9, n. 9, p. e202996891-e202996891.
- [43] ECHEVERRIA, L. E.; MORILLO, C.A. American Trypanosomiasis (Chagas Disease). **Infectious Disease Clinics of North America**, 2019. [S.L.], v. 33, n. 1, p. 119-134.
- [44] CHAO, C.; LEONE, J. L.; VIGLIANO, C. A. Chagas disease: historic perspective. **Biochimica Et Biophysica Acta (Bba) - Molecular Basis of Disease**, 2020. [S.L.], v. 1866, n. 5, p. 165689.
- [45] GUHL, F.; RAMÍREZ, J. D. Poverty, Migration, and Chagas Disease. **Current Tropical Medicine Reports**, 2021. [S.L.], v. 8, n. 1, p. 52-58.
- [46] CORTES, S. *et al.* Potential of the natural products against leishmaniasis in Old World - a review of in-vitro studies. **Pathogens and Global Health**, 2020. v. 114, n. 4, p. 170–182.
- [47] MACALUSO G. *et al.* A Review on the Immunological Response against *Trypanosoma cruzi*. **Pathogens**, 2023. 12(2):282.
- [48] DAUDA, M. *et al.* Journal of Infection and Public Health Scientometric analysis of global trypanosomiasis research: 1988 – 2017. **Journal of Infection and Public Health**, 2020. v. 13, n. 4, p. 514–520.
- [49] ORTIZ, J. E. *et al.* Candimine from *Hippeastrum escoipense* (Amaryllidaceae): anti-trypanosoma cruzi activity and synergistic effect with benznidazole. **Phytomedicine**, 2023. [S.L.], v. 114, p. 154788.
- [50] COSTA-OLIVEIRA, C. N. *et al.* Outbreak of Chagas disease in Brazil: validation of a molecular diagnostic method. **Experimental Parasitology**, 2023. [S.L.], v. 247, p. 108478.
- [51] CORREIA, J. R. *et al.* Doença de Chagas: aspectos clínicos, epidemiológicos e fisiopatológicos. **Revista Eletrônica Acervo Saúde**, 2021. v. 13, n. 3, p. e6502-e6502.
- [52] SANTOS E. F. *et al.* Acute chagas disease in brazil from 2001 to 2018: A nationwide spatiotemporal analysis. **PLoS Neglected Tropical Disease**, 2020; 14(8): 1–16.
- [53] XIMENES, M. F. F. M. Atlantic Forest and Caatinga: two threatened tropical biomes in Brazil and implications for public health. **Journal Of Biomedical Research & Environmental Sciences**, 2022. [S.L.], v. 3, n. 5, p. 641-658.
- [54] GARCÍA-HUERTAS, P.; CARDONA-CASTRO, N. Advances in the treatment of Chagas disease: promising new drugs, plants and targets. **Biomedicine & Pharmacotherapy**, 2021. [S.L.], v. 142, p. 112020.
- [55] FONSECA B. P. *et al.* Neglected tropical diseases in Brazil: lack of correlation between disease burden, research funding and output. **Tropical Medicine and International Health**, 2020; 25(2):1373-1384.

- [56] JAIN, C.; KHATANA, S.; VIJAYVERGIA, R. Bioactivity of Secondary Metabolites of Various Plants: A Review. **International Journal of Pharmaceutical Sciences and Research: International Journal of Pharmaceutical Sciences and Research**, 2019. [S.L.], v. 10, n. 2, p. 494-504.
- [57] ADJEI, M. O. *et al.* MicroRNAs Roles in Plants Secondary Metabolism. **Plant Signaling & Behavior**, 2021. [S.L.], v. 16, n. 7, p. 1915590.
- [58] BHUSHAN, S. *et al.* Recent Strategies to Engineer Alkaloid Biosynthesis in Medicinal Plants. **Medicinal Plants**, 2023. [S.L.], p. 391-416.
- [59] LIU, B. *et al.* Advances of Veratrum nigrum L. Steroid Alkaloids. **Industrial Crops and Products**, 2013. [S.L.], v. 191, p. 115946.
- [60] WANG, S. W. *et al.* Chemical Constituents and Antidepressant-Like Activity of the Ethanol Extract of *Lindera fragrans* Leaves. **Pharm Chem J**, 2023. 56, 1358–1368.
- [61] CHO, KIN-HAU. *et al.* Morphinan Alkaloids from the Leaves of *Alphonsea cylindrica* and Their Antibacterial Properties. **Planta Medica**, 2023. v. 89, n. 01, p. 79-85.
- [62] DENG M.Y.W; GU Z, L.I.Y.; CHEN L, H.E.B. Anti-Neuroinflammatory Potential of Natural Products in the Treatment of Alzheimer's Disease. **Molecules**. 2023; 28(3):1486.
- [63] TANG, S. *et al.* *Ephedrae Herba*: A Review of Its Phytochemistry, Pharmacology, Clinical Application, and Alkaloid Toxicity. **Molecules**, 2023. v. 28, n. 2, p. 663.
- [64] TANVEER, M. A. *et al.* Trigonelline, a plant derived alkaloid prevents ultraviolet-B-induced oxidative DNA damage in primary human dermal fibroblasts and BALB/c mice via modulation of phosphoinositide 3-kinase-Akt-Nrf2 signalling axis. **Experimental Gerontology**, 2023. v. 171, p. 112028.
- [65] AKINBOYE, A. J. *et al.* Alkaloids in food: a review of toxicity, analytical methods, occurrence and risk assessments. **Food Science and Biotechnology**, 2023. p. 1-26.
- [66] WALLER G. R.; NOWACKI E.K. The role of alkaloids in plants. In: *Alkaloid biology and metabolism in plants*. Springer: Boston, 1978. 143-181.
- [67] RAJAK, P. *et al.* Agricultural pesticides—friends or foes to biosphere?. **Journal of Hazardous Materials Advances**, 2023. v. 10, p. 100264.
- [68] ALSWAT, A. M. *et al.* Investigating the metagenomics of the bacterial communities in the rhizosphere of the desert plant *Senna italica* and their role as plant growth promoting factors. **Notulae Botanicae Horti Agrobotanici Cluj-Napoca**, 2023. v. 51, n. 1, p. 13053-13053.
- [69] WRÓŃSKA, A. K.; BOGUŚ, M. I. Harman and norharman, metabolites of the entomopathogenic fungus *Conidiobolus coronatus* (Entomophthorales), affect the serotonin levels and phagocytic activity of hemocytes, insect immunocompetent cells, in *Galleria mellonella* (Lepidoptera). **Cell BioSci**. 2019, 9, 29.
- [70] TAYAMA, Y. *et al.* Anti-Trypanosoma cruzi activity of Coptis rhizome extract and its constituents. **Tropical Medicine and Health**, 2023. v. 51, n. 1, p. 12.

- [71] GUTIÉRREZ-GRIJALVA, E. P. *et al.* Plant alkaloids: structures and bioactive properties. **Plant-Derived Bioactives: Chemistry and Mode of Action**, 2020. p. 85-117.
- [72] ADIBAH, K. Z. M.; AZZREENA, M. A. Plant toxins: alkaloids and their toxicities. **GSC Biological and Pharmaceutical Sciences**, 2019. v. 6, n. 2.
- [73] THAWABTEH, A. *et al.* The biological activity of natural alkaloids against herbivores, cancerous cells and pathogens. **Toxins**, 2019. v. 11, n. 11, p. 656.
- [74] LIN, X.; LI, X.. A review on applications of computational methods in drug screening and design. **Molecules**, 2020. v. 25, n. 6, p. 1375.
- [75] D'SOUZA, S.; PREMA, KV; BALAJI, S. Machine learning models for drug–target interactions: current knowledge and future directions. **Drug Discovery Today**, 2020. v. 25, n. 4, pág. 748-756.
- [76] GHISLAT, G.; RAHMAN, T.; BALLESTER, P. J. Recent progress on the prospective application of machine learning to structure-based virtual screening. **Current opinion in chemical biology**, 2021. v. 65, p. 28-34.
- [77] GORGULLA, C. *et al.* An open-source drug discovery platform enables ultra-large virtual screens. **Nature**, 2020. v. 580, n. 7805, p. 663-668.
- [78] DAGUR, P. *et al.* Virtual screening of phytochemicals for drug discovery. In: *Phytochemistry, Computational Tools and Databases in Drug Discovery*. **Elsevier**, 2023. p. 149-179.
- [79] SVENSSON, F.; KARLÉN, A.; SKÖLD, C. “Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods”. **Journal of Chemical Information and Modeling**, 2012. v. 52, p. 225–232.
- [80] ASHTON, T. D. *et al.* The Development Process for Discovery and Clinical Advancement of Modern Antimalarials. **Journal Of Medicinal Chemistry**, 2019. [S.L.], v. 62, n. 23, p. 10526-10562.
- [81] LIU, C. *et al.* Pnictogens in medicinal chemistry: evolution from erstwhile drugs to emerging layered photonic nanomedicine. **Chemical Society Reviews**, 2021. [S.L.], v. 50, n. 4, p. 2260-2279.
- [82] YASUO, N.; SEKIJIMA, M. Improved method of structure-based virtual screening via interaction-energy-based learning. **Journal of Chemical Information and Modeling**, 2019. v. 59, n. 3, p. 1050-1061.
- [83] GUPTA, R. *et al.* Inteligência artificial para aprendizado profundo: abordagem de inteligência de máquina para descoberta de medicamentos. **Diversidade molecular**, 2021. v. 25, p. 1315-1360.
- [84] HARIYONO, P. *et al.* An integrated virtual screening of compounds from Carica papaya leaves against multiple protein targets of SARS-Coronavirus-2. **Results in chemistry**, 2021. v. 3, p. 100113.

- [85] MEDEIROS, A. R. *et al.* Andricopulo, A.D. Chemoinformatics Studies on a Series of Imidazoles as Cruzain Inhibitors. **Biomolecules** 2021, 11, 579.
- [86] PIRES, D. E. V. *et al.* “EasyVS: A User-Friendly Web-Based Tool for Molecule Library Selection and Structure-Based Virtual Screening”. **Bioinformatics**, vol. 36, n. 14, julho de 2020, p. 4200–02.
- [87] ELGAMAL, A. M. *et al.* Mansour. Phytochemical profiling and anti-aging activities of Euphorbia retusa extract: in silico and in vitro studies. **Arabian Journal Of Chemistry**, 2021. [S.L.], v. 14, n. 6, p. 103159.
- [88] GIGLIONI, V. *et al.* The use of receiver operating characteristic curves and precision-versus-recall curves as performance metrics in unsupervised structural damage classification under changing environment. **Engineering Structures**, 2021. v. 246, p. 113029.
- [89] FIELDING, A. H.; BELL, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. **Environmental conservation**, 1997 v. 24, n. 1, p. 38-49.
- [90] POLO, T. C. F.; MIOT, H.A. Use of ROC curves in clinical and experimental studies. **J Vasc Bras**. 2020, 11-19.
- [91] CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, 2020. v.21, pp: 6-1–6- 13.
- [92] HANLEY, J. A. Receiver operating characteristic (ROC) methodology: the state of the art. **Crit Rev Diagn Imaging**. 1989. 29(3):307-35.
- [93] MOLYNEUX, D. H; SAVIOLI, L; ENGELS, D. Neglected tropical diseases: progress towards addressing the chronic pandemic. **Lancet**, 2017. 389(10066):312–25.
- [94] CASTALDO, M. *et al.* Anthropological study on Chagas Disease: Sociocultural construction of illness and embodiment of health barriers in Bolivian migrants in Rome, Italy. **Plos one**, 2020. v. 15, n. 10, p. e0240831.
- [95] OMS. Organização Mundial da Saúde. Neglected tropical diseases [Internet]. Geneva: World Health Organization; 2020. Disponível em: <https://www.who.int/neglected_diseases/diseases/en/>. Acesso em: 23/03/2023.
- [96] LENK, E. J. *et al.* Socioeconomic benefit to individuals of achieving 2020 targets for four neglected tropical diseases controlled/eliminated by innovative and intensified disease management: Human African trypanosomiasis, leprosy, visceral leishmaniasis, Chagas disease. **PLoS Negl. Trop. Dis**. 2018, 12, e0006250.
- [97] SANTOS, F. *et al.* Eco-epidemiology of vectorial *Trypanosoma cruzi* transmission in a region of northeast Brazil. **Acta Tropica**, 2022. v. 225, p. 106184.
- [98] VERMELHO, A. B.; RODRIGUES, G. C.; SUPURAN, C. T. Why hasn't there been more progress in new Chagas disease drug discovery?. **Expert Opinion on Drug Discovery**, 2020. v. 15, n. 2, p. 145-158.

- [99] AGRAWAL, P. *et al.* Benchmarking of different molecular docking methods for protein-peptide docking. **BMC Bioinformatics**, 2019. v. 19, n. S13, p. 426, 4.
- [100] PINZI, L.; RASTELLI, G. Molecular Docking: Shifting Paradigms in Drug Discovery. **International Journal of Molecular Sciences**, 2019. v. 20, n. 18, p. 4331, 4.
- [101] LENGAUER T, RAREY M. "Computational methods for biomolecular “docking””. **Curr. Opin. Struct. Biol**, 1996. V. 6, n. 3, p. 402–406.
- [102] CRAMPON, K. *et al.* Machine-learning methods for ligand–protein molecular docking. **Drug discovery today**, 2022. v. 27, n. 1, p. 151-164.
- [103] YU, Q.; PAVOŠEVIĆ, F.; HAMMES-SCHIFFER, S. Development of nuclear basis sets for multicomponent quantum chemistry methods. **The Journal of Chemical Physics**, 2020. v. 152, n. 24, p. 244123.
- [104] MOUCLIS, V. D. *et al.* Advances in de novo drug design: from conventional to machine learning methods. **International journal of molecular sciences**, 2021. v. 22, n. 4, p. 1676.
- [105] BATISTA, V. *et al.* Construção, Otimização e Ancoragem Molecular de Substâncias Bioativas em Biomacromoléculas: um tutorial prático. **Química Nova**, 2021. [S.L.], v. 45, p. 223-234.
- [106] MAIA, E. H. B. *et al.* Structure-based virtual screening: from classical to artificial intelligence. **Frontiers in chemistry**, 2020. v. 8, p. 343.
- [107] MARTÍNEZ, J. *et al.* Computational Studies of Aflatoxin B1 (AFB1): A Review. **Toxins**, 2023. v. 15, n. 2, p. 135.
- [108] KULSHRESTHA, S. *et al.* Advanced approaches and *in silico* tools of chemoinformatics in drug designing. In: **Chemoinformatics and Bioinformatics in the Pharmaceutical Sciences**. Academic Press, 2021. p. 173-206.
- [109] SIVAKUMAR, K. C. *et al.* Perspectivas de estratégias de design de drogas multialvo ligando docking molecular e dinâmica molecular para explorar o processo de reconhecimento de proteína-ligante. **Drug Development Research**, 2020. v. 81, n. 6, p. 685-699.
- [110] RAMASWAMY, A.; BALASUBRAMANIAN, S.; RAJAGOPALAN, M. Biomolecular talks—part 1: a theoretical revisit on molecular modeling and docking approaches. In: **Molecular Docking for Computer-Aided Drug Design**. Academic Press, 2021. p. 31-55.
- [111] SANTOS-SILVA, C. A. *et al.* Plant antimicrobial peptides: state of the art, *in silico* prediction and perspectives in the omics era. **Bioinformatics and Biology Insights**, 2020. v. 14, p. 1177932220952739.
- [112] SILAKARI, O.; SINGH, P. K. **Concepts and experimental protocols of modelling and informatics in drug design**. Academic Press, 2020. cap 6.
- [113] CHANDRASEKARAN, B. *et al.* Drug-Receptor Interactions. **Frontiers in Pharmacology of Neurotransmitters**, 2020. p. 31-68.

- [114] SMITH, J. M., VAN NESS, H. C., ABBOT, M. M. **Introdução à Termodinâmica da Engenharia Química**. Editora LTC, 7^a. Ed. 2007.
- [115] JI, B. *et al.* Prediction of the binding affinities and selectivity for CB1 and CB2 ligands using homology modeling, molecular docking, molecular dynamics simulations, and MM-PBSA binding free energy calculations. **ACS Chemical Neuroscience**, 2020. v. 11, n. 8, p. 1139-1158.
- [116] ADELUSI T. I. *et al.* Temitope Isaac *et al.* Molecular modeling in drug discovery. **Informatics in Medicine Unlocked**, 2022. [S.L.], v. 29, p. 100880.
- [117] LUISE, C. **Application of computer-based methods to identify and investigate ligands for proteins with high flexible binding sites**. 2022. Tese de Doutorado. Dissertation, Halle (Saale), Martin-Luther-Universität Halle-Wittenberg.
- [118] NYONGBELA, K. D.; NTIE-KANG, F.; HOYE, T. R.; EFANGE, S. M. N. Antiparasitic sesquiterpenes from the cameroonian spice *Scleria striatinux* and preliminary *in vitro* and *in silico* DMPK assessment. **Nat. Prod. Bioprospect.**, 2017, 7(3), 235-247.
- [119] RODRIGUES, D. F.; *et al.* Trypanocidal activity of Flavokawin B, a component of *Polygonum ferrugineum* Wedd. **Planta Med.**, 2017, 83(3-04), 239-244.
- [120] EZZAT, S. M.; *et al.* Antiprotozoal activity of major constituents from the bioactive fraction of *Verbesina encelioides*. **Nat. Prod. Res.**, 2017, 31(6), 676-680.
- [121] HEINRICH, M.; MAH, J.; AMIRKIA, V. Alkaloids used as medicines: Structural phytochemistry meets biodiversity—An update and forward look. **Molecules**, 2021. v. 26, n. 7, p. 1836.
- [122] YESHI, K.; CRAYN, D.; RITMEJERYTĚ, E.; WANGCHUK, P. Plant Secondary Metabolites Produced in Response to Abiotic Stresses Has Potential Application in Pharmaceutical Product Development. **Molecules**, 2022, 27, 313.
- [123] OLIVEIRA, C. V. B. *et al.* A Potential New Source of Therapeutic Agents for the Treatment of Mucocutaneous Leishmaniasis: the essential oil of *Rhaphidion echinus*. **Molecules**, 2022. v. 27, n. 7, p. 2169.
- [124] OZARDE, YOGITA S.; CHOUDHARI, VISHNU P. Preliminary Qualitative Phytochemical Analysis and Acute Oral Toxicity Study of Latex of an Ethnomedicinal Plant *Euphorbia fusiformis* Buch.-Ham. Ex D. Don. **Research Journal of Pharmacy and Technology**, 2022. [S.L.], p. 1123-1127.
- [125] CARVALHO, M. G. *et al.* *Croton pulegioides* Baill and *Croton piauhiensis* Mull. Arg. (Euphorbiaceae) Essential Oils: chemical composition and anti-leishmania activity. **Revista Virtual de Química**, 2022. [S.L.], v. 14, n. 6, p. 938-946.
- [126] NÚÑEZ, M. J. *et al.* *In vitro* antiparasitic and antibacterial evaluation of organic extracts of Salvadoran flora. **Bol Latinoam Caribe Plant Med Aromat**, 2023. 22(1): 19-36.
- [127] NAJMI, Asim *et al.* Modern approaches in the discovery and development of plant-based natural products and their analogues as potential therapeutic agents. **Molecules**, 2022. v. 27, n. 2, p. 349.

- [128] BARNES, E. C.; KUMAR, R.; DAVIS, R. A. The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. **Nat. Prod. Rep.**, 2016. 33, 372–381.
- [129] TINTORE, M.; VIDAL-JORDANA, A.; SASTRE-GARRIGA, J. Treatment of multiple sclerosis — success from bench to bedside. **Nat. Rev. Neurol.**, 2019. 15, 53–58.
- [130] MOSHAWIH, S. *et al.* Synergy between machine learning and natural products cheminformatics: Application to the lead discovery of anthraquinone derivatives. **Chemical Biology & Drug Design**, 2022. v. 100, n. 2, p. 185-217.
- [131] LIPINSKI, C. A. *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Adv. Drug Delivery Rev.**, 1997, 23, 3– 25.
- [132] OWEN, A. E. *et al.* Natural Andrographolide Isolated from *Andrographis paniculata* as Potent Epileptic Agent: Spectroscopy, Molecular Structure, and Molecular Docking Investigation. **Chemistry Africa**, 2013. p. 1-17.
- [133] SHULTZ, M. D. Two decades under the influence of the rule of five and the changing properties of approved oral drugs. **J. Med. Chem.**, 2019. 62, 1701–1714.
- [134] GRIGALUNAS, M. *et al.* Produtos pseudonaturais e métodos inspirados em produtos naturais em biologia química e descoberta de medicamentos. **Current Opinion in Chemical Biology**, 2020. v. 56, p. 111-118.
- [135] MARCHEV, A. S. *et al.* Metabolomics and health: From nutritional crops and plant-based pharmaceuticals to profiling of human biofluids. **Cellular and Molecular Life Sciences**, 2021. v. 78, n. 19-20, p. 6487-6503.
- [136] SÜNTAR, I. Importance of ethnopharmacological studies in drug discovery: role of medicinal plants. **Phytochemistry Reviews**, 2020. v. 19, n. 5, p. 1199-1209.
- [137] BARBA-OSTRIA, C. *et al.* Evaluation of biological activity of natural compounds: Current trends and methods. **Molecules**, v. 27, n. 14, p. 4490, 2022.
- [138] LANGEDER, J. *et al.* Natural products against acute respiratory infections: Strategies and lessons learned. **Journal of ethnopharmacology**, 2020. v. 248, p. 112298.
- [139] CAESAR, L. K. *et al.* Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities. **Natural product reports**, 2021 v. 38, n. 11, p. 2041-2065.
- [140] BALACHANDRAN, P. Ayurvedic Knowledge Inspired Approach to Modern Drug Discovery. In: Chemistry, Biological Activities and Therapeutic Applications of Medicinal Plants in Ayurveda. **Royal Society of Chemistry**, 2022. p. 325-373.
- [141] ALMALKI, G. H. *et al.* Immunohistochemical evaluation of the *euphorbia inarticulata* extract on liver and kidney tissues in hepatocellular carcinoma rats. **Pharmacophore**, 2022. v. 13, n. 2, p. 33-40.

- [142] TEREFE, E. M. **In vitro and in silico pharmacologic evaluation of the antiretroviral activity of croton species**. 2022. Tese de Doutorado. University of Nairobi.
- [143] TANGSARIYA, J.; SHARMA, A. Pollen Morphological and Taxonomical Studies in Some Members of Family Euphorbiaceae. **Voyager**, 2021. v.7, n.2, p. 70-78.
- [144] MURAGURI, S.; LIU, A. Castor Bean: Recent Progress in Understanding the Genome of This Underutilized Crop. **Underutilised Crop Genomes**, 2022. p. 337-354.
- [145] KAUFMAN, S. R.; KAUFMAN, W. **Invasive plants: guide to identification and the impacts and control of common North American species**. Rowman & Littlefield, 2023.
- [146] WEBSTER G. L. Conspectus of a new classification of the Euphorbiaceae. **Taxon**, 1975. 24: 593-601.
- [147] CRAENE, L. P. R. **Floral diagrams: an aid to understanding flower morphology and evolution**. Cambridge University Press, 2022.
- [148] ANTINORI, S. *et al.* Family cluster of Chagas disease among Bolivian immigrants in Italy: High rate of maternal-fetal transmission. **Travel Medicine and Infectious Disease**, 2022. v. 49, p. 102370.
- [149] SHIKANAI YASUDA, M. A. Emerging and reemerging forms of Trypanosoma cruzi transmission. **Memórias do Instituto Oswaldo Cruz**, 2021. v. 117.
- [150] FLORES-CHAVEZ, M. D. *et al.* Parasitemia Levels in Trypanosoma cruzi Infection in Spain, an Area Where the Disease Is Not Endemic: Trends by Different Molecular Approaches. **Microbiology Spectrum**, 2022. v. 10, n. 5, p. e02628-22.
- [151] BRANCO SANTOS, J. C. *et al.* Bisphosphonate-based molecules as potential new antiparasitic drugs. **Molecules**, 2020. v. 25, n. 11, p. 2602.
- [152] SAAD, E. J. *et al.* Infections in the first year after renal transplant. **Revista Medicina**, 2020. v. 80, n. 6.
- [153] LA HOZ, R. M.; MORRIS, M. I. INFECTIOUS DISEASES COMMUNITY OF PRACTICE OF THE AMERICAN SOCIETY OF TRANSPLANTATION. Tissue and blood protozoa including toxoplasmosis, Chagas disease, leishmaniasis, Babesia, Acanthamoeba, Balamuthia, and Naegleria in solid organ transplant recipients—Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. **Clinical transplantation**, 2019. v. 33, n. 9, p. e13546.
- [154] RIARTE A. *et al.* Chagas disease in patients with kidney transplants: 7 years of experience 1989–1996. **Clin Infect Dis**, 1999. 29: 561–67.
- [155] MARTÍN-ESCOLANO, Javier *et al.* An updated view of the trypanosoma cruzi life cycle: Intervention points for an effective treatment. **ACS Infectious Diseases**, 2022. v. 8, n. 6, p. 1107-1115.
- [156] NGUMBI, ANNA FLOWINO. **Use and misuse of trypanocides and comparison of putatively Drug sensitive and resistant strains of trypanosome Congolense isolated from Tanzania**. 2020. Tese de Doutorado. Sokoine University of Agriculture.

- [157] ZUMA, A. A.; DOS SANTOS BARRIAS, E.; DE SOUZA, W. Basic biology of trypanosoma cruzi. **Current Pharmaceutical Design**, 2021. v. 27, n. 14, p. 1671-1732.
- [158] JIMÉNEZ, P. *et al.* A systematic review of the Trypanosoma cruzi genetic heterogeneity, host immune response and genetic factors as plausible drivers of chronic chagasic cardiomyopathy. **Parasitology**, 2019. v. 146, n. 3, p. 269-283.
- [159] AUSTEN, J. M.; BARBOSA, A. D. Diversity and epidemiology of bat trypanosomes: a one health perspective. **Pathogens**, 2021. v. 10, n. 9, p. 1148.
- [160] LEITE, A. L. J. **Regulatory activity of the extracellular degradation pathway of adenine nucleotides in the virulence of different strains and evolutionary forms of Trypanosoma cruzi in vitro and in vivo**. 2020. Tese de Doutorado. Federal University of Ouro Preto.
- [161] MALAVAZI, P. F. N. S *et al.* Trypanosomes of vectors and domestic dogs in Trypanosoma cruzi transmission areas from Brazilian southwestern amazon: New mammalian host for Trypanosoma janseni. **Acta Tropica**, 2020. v. 210, p. 105504.
- [162] BITTINELLI, I. F. *et al.* Do not judge a book by its cover: would Triatoma tibiamaculata (Pinto, 1926) belong to Triatoma Laporte, 1832, or to Panstrongylus Berg, 1879, with misleading homoplasies?. **Parasites & Vectors**, 2022. v. 15, n. 1, p. 184.
- [163] SANTOS-MIRANDA, Artur *et al.* Reactive oxygen species and nitric oxide imbalances lead to in vivo and in vitro arrhythmogenic phenotype in acute phase of experimental Chagas disease. **PLoS pathogens**, 2020. v. 16, n. 3, p. e1008379.
- [164] RADISIC, Marcelo V.; REPETTO, Silvia A. American trypanosomiasis (Chagas disease) in solid organ transplantation. **Transplant Infectious Disease**, 2020. v. 22, n. 6, p. e13429.
- [165] LEVIN, Donald A.; YORK JR, Billie M. The toxicity of plant alkaloids: an ecogeographic perspective. **Biochemical Systematics and Ecology**, 1978. v. 6, n. 1, p. 61-76.
- [166] ALTCHER, J. M. Congenital Chagas disease. **Chagas Disease: A Clinical Approach**, 2019. p. 179-198.
- [167] SAMPAIO, G.H.F. *et al.* Epidemiological profile of acute Chagas disease in individuals infected by oral transmission in northern Brazil. **Rev. Soc. Bras. Med. Trop.** 2020, 53, p.e20200088.
- [168] HASSLOCHER-MORENO, A. M. *et al.* Progression rate from the indeterminate form to the cardiac form in patients with chronic Chagas disease: twenty-two-year follow-up in a Brazilian urban cohort. **Tropical Medicine and Infectious Disease**, 2020. v. 5, n. 2, p. 76.
- [169] CHADALAWADA, Sindhu *et al.* Risk of chronic cardiomyopathy among patients with the acute phase or indeterminate form of Chagas disease: a systematic review and meta-analysis. **JAMA network open**, 2020. v. 3, n. 8, p. e2015072-e2015072.
- [170] WARD, A. I. *et al.* Trypanosoma cruzi amastigotes that persist in the colon during chronic stage murine infections have a reduced replication rate. **Open biology**, 2020. v. 10, n. 12, p. 200261.

- [171] BALLINAS-VERDUGO, M. A. *et al.* Circulating miR-146a as a possible candidate biomarker in the indeterminate phase of Chagas disease. **Biological Research**, 2021. v. 54, n. 1, p. 1-16.
- [172] SERENO, D. *et al.* Noninvasive biological samples to detect and diagnose infections due to trypanosomatidae parasites: A systematic review and meta-analysis. **International Journal of Molecular Sciences**, 2020. v. 21, n. 5, p. 1684.
- [173] PEVERENGO, L. M. *et al.* Congenital chagas disease: Development and assessment of a specific IgM capture-based assay for diagnosis of transmission. **Acta Tropica**, 2021. v. 213, p. 105738.
- [174] THAKUR, S.; JOSHI, J; KAUR, S. Leishmaniasis diagnosis: an update on the use of parasitological, immunological and molecular methods. **Journal of Parasitic Diseases**, 2020. v. 44, p. 253-272.
- [175] CHATELAIN, E.; SCANDALE, I. Animal models of Chagas disease and their translational value to drug development. **Expert Opinion on Drug Discovery**, 2020. v. 15, n. 12, p. 1381-1402.
- [176] MANSOLDO, F. R. P. *et al.* Chagas disease: perspectives on the past and present and challenges in drug discovery. **Molecules**, 2020. v. 25, n. 22, p. 5483.
- [177] TWAIJ, B. M.; HASAN, M. N. Bioactive secondary metabolites from plant sources: Types, synthesis, and their therapeutic uses. **International Journal of Plant Biology**, 2022. v. 13, n. 1, p. 4-14.
- [178] FAISAL, S. *et al.* Alkaloids as potential antivirals. A comprehensive review. **Natural Products and Bioprospecting**, 2023. v. 13, n. 1, p. 4.
- [179] GAVHALE, S. S.; WAGHMARE, S. H A.; KAMBLE, H. V. A Review on: Alkaloids Important Therapeutic Compound. **World Journal of Pharmaceutical Research**, 2023. v. 12, n. 5. p. 741-752.
- [180] GIANNENAS, I. *et al.* The history of herbs, medicinal and aromatic plants, and their extracts: Past, current situation and future perspectives. In: **Feed additives**. Academic Press, 2020. p. 1-18.
- [181] NUNES, Clara dos Reis *et al.* Plants as sources of anti-inflammatory agents. **Molecules**, 2020. v. 25, n. 16, p. 3726.
- [182] Daylight Chemical Information Systems, Inc. **Daylight Theory Manual**. Califórnia, EUA 2011. Disponível em: <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>. Acesso em: maio de 2023.
- [183] TORTORELLA, S. *et al.* Combining machine learning and quantum mechanics yields more chemically aware molecular descriptors for medicinal chemistry applications. **Journal of Computational Chemistry**, 2021. v. 42, n. 29, p. 2068-2078.
- [184] CRUCIANI, G.; PASTOR, M.; GUBA, W. VolSurf: A New Tool for the Pharmacokinetic Optimization of Lead Compounds. **Eur. J. Pharm. Sci.**, 2000, 11: 29-39.

- [185] CRIVORI, P. *et al.* Predicting Blood-Brain Barrier Permeation from Three-Dimensional Molecular Structure. **J. Med. Chem.**, 2000, 43, 2204–2216.
- [186] CHEN, Y. *et al.* Comparison of dimension reduction methods on fatty acids food source study. **Scientific reports**, 2021. v. 11, n. 1, p. 1–12.
- [187] SYDOW, D. *et al.* TeachOpenCADD-KNIME: A Teaching Platform for Computer-Aided Drug Design Using KNIME Workflows. **Journal of Chemical Information and Modeling**, 2019. p. 8–11.
- [188] DIETZ, C. *et al.* Integration of the ImageJ Ecosystem in KNIME Analytics Platform. *Frontiers in Computer Science*, 2020. v. 2, p. 1–17.
- [189] KIM, YG.; MOON, K.IL. Clustering of smart meter big data based on KNIME analytic platform. **The Journal of the Institute of Internet, Broadcasting and Communication**, 2020. v. 20, n. 2, p. 13-20.
- [190] LÖFSTRÖM, T.; RYASIK, A.; JOHANSSON, Ulf. Tutorial for using conformal prediction in KNIME. In: **Conformal and Probabilistic Prediction with Applications**. PMLR, 2022. p. 4-23.
- [191] ALVES, V. M.; BRAGA, R. C.; MURATOV, E. N.; ANDRADE, C. H. Cheminformatics: An introduction. **Química Nova**, 2018. v. 41, n. 2, p. 202–212.
- [192] LIAO, L. *et al.* A Supervoxel-Based Random Forest Method for Robust and Effective Airborne LiDAR Point Cloud Classification. **Remote Sensing**, 2022. v. 14, n. 6, p. 1516.
- [193] IKRAM, S. T. *et al.* Prediction of IIoT traffic using a modified whale optimization approach integrated with random forest classifier. **The Journal of Supercomputing**, 2022. v. 78, n. 8, p. 10725-10756.
- [194] MOHAMMADI, S. *et al.* Ensemble learning from ensemble docking: Revisiting the optimum ensemble size problem. **Scientific Reports**, 2022. v. 12, n. 1, p. 1-15.
- [195] AI, Dongmei *et al.* Using decision tree aggregation with random forest model to identify gut microbes associated with colorectal cancer. **Genes**, 2022. v. 10, n. 2, p. 112.
- [196] DING, X-F. *et al.* Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: a secondary analysis of a cohort study. **Journal of translational medicine**, 2019. v. 17, n. 1, p. 1-10.
- [197] LI, X. *et al.* Combining machine learning and high-throughput experimentation to discover photocatalytically active organic molecules. **Chemical Science**, 2021 v. 12, n. 32, p. 10742-10754.
- [198] RODRÍGUEZ-PÉREZ, R.; MILJKOVIĆ, F.; BAJORATH, J. Assessing the information content of structural and protein–ligand interaction representations for the classification of kinase inhibitor binding modes via machine learning and active learning. **Journal of Cheminformatics**, 2020. v. 12, n. 1, p. 1-14.

- [199] GOEL, M. *et al.* Efficient and enhanced sampling of drug-like chemical space for virtual screening and molecular design using modern machine learning methods. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, 2023. v. 13, n. 2, p. e1637.
- [200] CHEN, B. **Molecular Graph Representation Learning and Generation for Drug Discovery**. 2022. Tese de Doutorado. Massachusetts Institute of Technology.
- [201] LEMM, D.; VON RUDORFF, G. F.; VON LILIENFELD, O. Anatole. Machine learning based energy-free structure predictions of molecules, transition states, and solids. **Nature Communications**, 2021. v. 12, n. 1, p. 4468.
- [202] SINGH, S.; JALAL, M.; KAUR, S. Bug Classification Depend Upon Refactoring Area of Code. **Journal of The Institution of Engineers (India): Series B**, 2023. p. 1-17.
- [203] GULTEPE, Y; RASHED, S. The use of data mining techniques in heart disease prediction. **Int. J. Comput. Sci. Mob. Comput**, 2019. v. 8, n. 4, p. 136-141.
- [204] MAZANETZ, M. P.; GOODE, C. H. F; CHUDYK, E. I. Ligand-and structure-based drug design and optimization using KNIME. **Current Medicinal Chemistry**, 2020. v. 27, n. 38, p. 6458-6479.
- [205] VETOVA, S. Workflow model for big data analysis and visualization. **International Journal of Applied Mathematics, Computational Science and Systems Engineering**, 2021. v. 3, p. 99-106.
- [207] ANDERSON, A. W. *et al.* Independent external validation of artificial intelligence algorithms for automated interpretation of screening mammography: a systematic review. **Journal of the American College of Radiology**, 2022. v. 19, n. 2, p. 259-273.
- [208] CZUB, N. *et al.* Artificial Intelligence-Based Quantitative Structure–Property Relationship Model for Predicting Human Intestinal Absorption of Compounds with Serotonergic Activity. **Molecular Pharmaceutics**, 2023. v. 20, n. 5, p. 2545-2555.
- [209] ZENG, B. *et al.* A correlation analysis of short-term imaging manifestations and long-term function using ROC curve after tibial fracture surgery. **American Journal of Translation Research**, 2021. v. 13, n. 6, p. 6724-6730.
- [210] CHICCO, D.; JURMAN, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. **BioData Mining**, 2023. v. 16, n. 1, p. 1-23.
- [211] HARRIS, L. *et al.* A generative deep learning approach to stochastic downscaling of precipitation forecasts. **Journal of Advances in Modeling Earth Systems**, 2022. v. 14, n. 10, p. e2022MS003120.
- [212] ZHU, R. *et al.* IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier. **BMC bioinformatics**. 2021. v. 22, n.1, p.175.
- [213] SARAVANAMUTHU, A. *et al.* Two Interacting Binding Sites for Quinacrine Derivatives in the Active Site of Trypanothione Reductase. **Journal Of Biological Chemistry**, 2004. v. 279, n. 28, p. 29493-29500.

- [214] SCHORMANN, N. *et al.* Crystal structure of Trypanosoma cruzi pteridine reductase 2 in complex with a substrate and an inhibitor. **Journal of structural biology**, 2005. v. 152, n. 1, p. 64-75.
- [215] SENKOVICH, O.; SCHORMANN, N.; CHATTOPADHYAY, D. Structures of dihydrofolate reductase-thymidylate synthase of Trypanosoma cruzi in the folate-free state and in complex with two antifolate drugs, trimetrexate and methotrexate. **Acta Crystallographica Section D: Biological Crystallography**, 2009. v. 65, n. 7, p. 704-716.
- [216] CHEN, Y. T. *et al.* In vitro and in vivo studies of the trypanocidal properties of WRR-483 against Trypanosoma cruzi. **PLoS neglected tropical diseases**, 2010. v. 4, n. 9, p. e825.
- [217] BITENCOURT-FERREIRA, G.; DE AZEVEDO, W. F. Molegro virtual docker for docking. **Docking screens for drug discovery**, 2019. p. 149-167.
- [218] ESCAMILLA-GUTIÉRREZ, A. *et al.* In silico selection of aptamers for bacterial toxins detection. **Journal of Biomolecular Structure and Dynamics**, 2022. p. 1-10.
- [219] PADHORNY, D. *et al.* Protein–ligand docking using FFT based sampling: D3R case study. **Journal of Computer-Aided Molecular Design**, 2018. v. 32, n. 1, p. 225–230.
- [220] ABRAHAM, M. J. *et al.* GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. **SoftwareX**, 2015, 1, 19–25.
- [221] BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. **Comput. Phys. Commun.** 1995, 91 (1–3), 43–56.
- [222] MALDE, A. K.; *et al.* An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. **J. Chem. Theory Comput.** 2011, 7 (12), 4026–4037.
- [223] BONDI, A. Van Der Waals Volumes and Radii. **J. Phys. Chem.** 1964, 68 (3), 441–451.
- [224] Pettersen, E. F. *et al.* UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. **Protein Sci.** 2021, 30 (1), 70–82.
- [225] PIRES D. E. V.; BLUNDELL T. L.; ASCHER D. B. pkCSM: predicting smallmolecule pharmacokinetic and toxicity properties using graph-based signatures, **J Med Chem**, 2018; 58(9):4066-4072.
- [226] MAHATO S. *et al.* Synthesis, *in silico* studies and in vitro evaluation for antioxidant and antibacterial properties of diarylmethylamines: a novel class of structurally simple and highly potent pharmacophore. **Eur J Pharm Sci.** 2016. v.88. p. 88202-209.
- [227] CEPERLEY, D. M.; LIBBY, S. B. Berni Julian Alder, theoretical physicist and inventor of molecular dynamics, 1925–2020. **Proceedings of the National Academy of Sciences**, 2021. v. 118, n. 11, p. e2024252118.
- [228] ALDER B. J.; WAINWRIGHT T. E. Phase Transition for a Hard Sphere System. **J Chem Phys**, 1957. 27, 1208–1209.

- [229] TSIMPANOIANNIS, I. N. *et al.* Self-diffusion coefficient of bulk and confined water: a critical review of classical molecular simulation studies. **Molecular Simulation**, 2019. v. 45, n. 4-5, p. 425-453.
- [230] CHEN, G. *et al.* Molecular dynamics simulation for mechanism elucidation of food processing and safety: State of the art. **Comprehensive reviews in food science and food safety**, 2019. v. 18, n. 1, p. 243-263.
- [231] WU, J. *et al.* Application of molecular dynamics simulation for exploring the roles of plant biomolecules in promoting environmental health. **Science of The Total Environment**, 2023. v. 869, p. 161871.
- [232] LOKHANDE, K. B. *et al.* High throughput virtual screening and molecular dynamics simulation analysis of phytomolecules against BfmR of *Acinetobacter baumannii*: anti-virulent drug development campaign. **Journal of Biomolecular Structure and Dynamics**, 2022. p. 1-15.
- [233] PEREZ, S.; MAKSHAKOVA, O. Multifaceted Computational Modeling in Glycoscience. **Chemical Reviews**, 2022. v. 122, n. 20, p. 15914-15970.
- [234] SUPLATOV, D.; SHARAPOVA, Y.; ŠVEDAS, V. EasyAmber: A comprehensive toolbox to automate the molecular dynamics simulation of proteins. **Journal of Bioinformatics and Computational Biology**, 2020. v. 18, n. 06, p. 2040011.
- [235] LIU, X. *et al.* Molecular Dynamics Simulations and Novel Drug Discovery. **Expert Opin. Drug Discov.**, 2018, 13, 23–37.
- [236] COURNIA, Z. *et al.* Rigorous free energy simulations in virtual screening. **Journal of chemical information and modeling**, 2020. v. 60, n. 9, p. 4153-4169.
- [237] FAHEEM, A. B. *et al.* Efficient parameterization of intermolecular force fields for molecular dynamics simulations via genetic algorithms. **Journal of Molecular Liquids**, 2021. v. 337, p. 116579.
- [238] KRISHNA, S.; SREEDHAR, I.; PATEL, C. M. Molecular dynamics simulation of polyamide-based materials—A review. **Computational Materials Science**, 2021. v. 200, p. 110853.
- [239] JONES, D. *et al.* Accelerators for classical molecular dynamics simulations of biomolecules. **Journal of chemical theory and computation**, 2022. v. 18, n. 7, p. 4047-4069.
- [240] SUH, D. *et al.* CHARMM-GUI Enhanced Sampler for molecular dynamics simulations with collective variables and various enhanced sampling methods. **Biophysical Journal**, 2022. v. 121, n. 3, p. 274a.
- [241] KAR, Supratik; ROY, Kunal; LESZCZYNSKI, Jerzy. *In silico* tools and software to predict ADMET of new drug candidates. In: **in silico methods for predicting drug toxicity**. New York, NY: Springer US, 2022. p. 85-115.
- [242] ALI, Yasir *et al.* Fragment-based approaches identified tecovirimat-competitive novel drug candidate for targeting the F13 protein of the monkeypox virus. **Viruses**, 2023. v. 15, n. 2, p. 570.

- [243] KAR, S.; ROY, K.; LESZCZYNSKI, J. *In silico* tools and software to predict ADMET of new drug candidates. In: ***in silico* methods for predicting drug toxicity**. New York, NY: Springer US, 2022. p. 85-115.
- [244] ALI, Y. et al. Fragment-based approaches identified tecovirimat-competitive novel drug candidate for targeting the F13 protein of the monkeypox virus. **Viruses**, 2023. v. 15, n. 2, p. 570.
- [255] SHELAKI, R. M.; PRAMANIK, D.; KIM, J-Y. Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. **Plant Biotechnology Reports**, 2019. v. 13, p. 423-445.
- [256] MELI, R. et al. Oxidative stress and BPA toxicity: an antioxidant approach for male and female reproductive dysfunction. **Antioxidants**, 2020. v. 9, n. 5, p. 405.
- [257] LI, Y.; LI, L. Contact dermatitis: classifications and management. **Clinical Reviews in Allergy & Immunology**, 2021. v. 61, n. 3, p. 245-281.
- [258] SUN, D. et al. Why 90% of clinical drug development fails and how to improve it?. **Acta Pharmaceutica Sinica B**, 2022. v. 12, n. 7, p. 3049-3062.
- [259] HAMED, R. et al. *In silico* screening for clinical efficacy of antiseizure medications: not all central nervous system drugs are alike. **Epilepsia**, 2023. v. 64, n. 2, p. 311.
- [260] RZEPIELA, A. A. et al. Conformational Effects on the Passive Membrane Permeability of Synthetic Macrocycles. **Journal of medicinal chemistry**, 2022. v. 65, n. 15, p. 10300-10317.
- [261] FEDI, A. et al. In vitro models replicating the human intestinal epithelium for absorption and metabolism studies: A systematic review. **Journal of Controlled Release** **335**. 2021. 247-268.
- [262] JOHNSON, Erik K. et al. Test negative designs with uncertainty, sensitivity, and specificity. **medRxiv**, 2021. p. 2021.06. 24.21259495.
- [263] BREDEWOUT, Mirte. **Virtual screening in multi-target drug design**. 32 f. Dissertação (Mestrado) - Department of Pharmaceutical Sciences, Utrecht University, 2022.
- [264] NACHBAGAUER, R. *et al.* A Chimeric Hemagglutinin-Based Universal Influenza Virus Vaccine Approach Induces Broad and Long-Lasting Immunity in a Randomized, Placebo-Controlled Phase I Trial. **Nat. Med.** 2021, 27 (1), 106–114.
- [265] SENKOVICH, O.; SCHORMANN, N.; CHATTOPADHYAY, D. Structures of Dihydrofolate Reductase-Thymidylate Synthase of Trypanosoma Cruzi in the Folate-Free State and in Complex with Two Antifolate Drugs, Trimetrexate and Methotrexate. **Acta Crystallogr. Sect. D Biol. Crystallogr.** 2009, 65 (7), 704–716.

ANEXOS

ANEXO 1

Tabela 1. Código SMILES das moléculas do banco de dados criado de alcaloides isolados da família Euphorbiaceae.

ID	Nome do composto	SMILES
1	Maytansine	<chem>CC1C2CC(C(C=CC=C(CC3=CC(=C(C(=C3)OC)C1)N(C(=O)CC(C4(C1O4)C)OC(=O)C(C)N(C)C(=O)C)C)OC)(NC(=O)O2)O</chem>
2	Parvistemonine	<chem>CC1CC(OC1=O)C2CCC3N2CCCCC3C4C(OC5C4OC(=O)C5C)C</chem>
3	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
4	Psycholeine	<chem>CC1CNC2C34C1(C(NC5=C(C=CC=C53)C67CCN(C6NC8=CC=CC=C78)C)N(CC4)C)C9=CC=CC(=C9N2)C12CCN(C1NC1=CC=CC=C21)C</chem>
5	Rhazinilam	<chem>CCC12CCCN3C1=C(C=C3)C4=CC=CC=C4NC(=O)CC2</chem>
6	Leuconolam	<chem>CCC12CCCN3C1(C(=CC3=O)C4=CC=CC=C4NC(=O)CC2)O</chem>
7	Viroallosecurinine	<chem>C1CCN2[C@H](C1)[C@@]34C[C@@H]2C=CC3=CC(=O)O4</chem>
8	Phyllantidine	<chem>C1CCN2[C@@H](C1)[C@]34C[C@H](O2)C=CC3=CC(=O)O4</chem>
9	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
10	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
11	Phyllantidine	<chem>C1CCN2[C@@H](C1)[C@]34C[C@H](O2)C=CC3=CC(=O)O4</chem>
12	Aralionine, debenzoyl-	<chem>CCC(C)C(C(=O)NC1C(OC2=CC=C(C=C2)/C=C/NC(=O)CNC1=O)C3=CC=CC=C3)N(C)C</chem>
13	Adouetine Y	<chem>CC[C@H](C)[C@H]1C(=O)N/C=C\C2=CC=C(C=C2)O[C@@H]([C@@H](C(=O)N1)NC(=O)[C@H](CC3=CC=CC=C3)N(C)C)C4=CC=CC=C4</chem>
14	Aristolochic acid C	<chem>C1OC2=C(O1)C3=C4C=C(C=CC4=CC(=C3C(=C2)C(=O)O)[N+](=O)[O-])O</chem>
15	Niruroidine	<chem>C1C[C@H]2[C@@]34C[C@@H]([C@@H](N2C1)CC3=CC(=O)O4)O</chem>
16	Isobubbialine	<chem>C1C[C@@H]2[C@@]34C[C@@H]([C@@H](N2C1)CC3=CC(=O)O4)O</chem>
17	Epibubbialine	<chem>C1C[C@H]2[C@@]34C[C@H]([C@@H](N2C1)CC3=CC(=O)O4)O</chem>

18	Bisbenzylisoquinoline ether	<chem>C1=CC=C(C=C1)CC2=C3C=CC=C4C3=C(O4)C(=N2)CC5=CC=CC=C5</chem>
19	2-substituted quinolines	<chem>C1=CC=C2C(=C1)C=CC(=N2)CCO</chem>
20	Aporphine	<chem>CN1CCC2=C3C1CC4=CC=CC=C4C3=CC=C2</chem>
21	Glaucine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=CC(=C(C=C43)OC)OC)OC)OC</chem>
22	Oxoglauicine	<chem>COC1=C(C2=C3C(=C1)C=CN=C3C(=O)C4=CC(=C(C=C42)OC)OC)OC</chem>
23	Salutaridine	<chem>CN1CC[C@]23C=C(C(=O)C=C2[C@H]1CC4=C3C(=C(C=C4)OC)O)OC</chem>
24	tetrahydroprotoberberine	<chem>C1CN2CC3=CC=CC=C3CC2C4=CC=CC=C41</chem>
25	HEMIARGYRINE	<chem>c1(c(cc2c(c1)C[C@H]1N([C@H]2C)CCc2c1cc(c(c2)OC)O)OC)O</chem>
26	Malloapeltine	<chem>COC1=C(C=[N+](C=C1)[O-])C#N</chem>
27	Bukittinggine	<chem>CC1CN2[C@H]3[C@H]4CCC35CCC[C@H]5[C@@]67C2[C@@H]1CC[C@@]46COC(=O)CC7</chem>
28	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
29	Phyllochryisine	<chem>C1CCN2C(C1)C34CC2C=CC3=CC(=O)O4</chem>
30	Phyllanthine	<chem>CO[C@@H]1CCN2[C@H](C1)C34C[C@H]2C=CC3=CC(=O)O4</chem>
31	Phyllalbine	<chem>CN1C2CCC1CC(C2)OC(=O)C3=CC(=C(C=C3)O)OC</chem>
32	Phyllantidine	<chem>C1CCN2[C@@H](C1)[C@]34C[C@H](O2)C=CC3=CC(=O)O4</chem>
33	Securinine	<chem>C1CCN2[C@H](C1)[C@]34C[C@H]2C=CC3=CC(=O)O4</chem>
34	Isostemonamine	<chem>CC1=C2CCCCN3C2(CCC3)C4(C1=O)C(=C(C(=O)O4)C)OC</chem>
35	Neostemonine	<chem>C[C@H]\1[C@@H]2[C@@H]3CCCN3CCC[C@H]2O/C1=C/4\C(=C(C(=O)O4)C)OC</chem>
36	Bisdehydroneostemonine	<chem>C[C@H]\1[C@H]2[C@@H](CCCN3C2=CC=C3)O/C1=C/4\C(=C(C(=O)O4)C)OC</chem>
37	Bisdehydroprotostemonine	<chem>C[C@H]1C[C@H](OC1=O)C2=CC=C3N2CCC[C@@H]4[C@@H]3[C@@H](/C(=C\5/C(=C(C(=O)O5)C)OC)/O4)C</chem>
38	Isoprotostemonine	<chem>C[C@H]1C[C@H](OC1=O)[C@@H]2CC[C@@H]3N2CCC[C@@H]4[C@@H]3[C@@H](/C(=C\5/C(=C(C(=O)O5)C)OC)/O4)C</chem>
39	Neotuberostemonine	<chem>CC[C@@H]1[C@H]2CCCCN3[C@H]2[C@@H](C[C@H]3[C@@H]4C[C@@H](C(=O)O4)C)[C@H]5[C@@H]1OC(=O)[C@H]5C</chem>
40	Bisdehydroneotuberostemonine	<chem>CCC1C2CCCCN3C2=C(C=C3C4CC(C(=O)O4)C)C5C1OC(=O)C5C</chem>

41	(E)-3-(3,5-Dimethoxy-4-hydroxybenzylidene)indoline-2-one	<chem>COC1=CC(=CC(=C1O)OC)/C=C\2/C3=CC=CC=C3NC2=O</chem>
42	2,3-Dihydro-1H-pyrrolo[2,1-c][1,4]benzodiazepine-5,11(10H,11ah)-dione	<chem>C1CC2C(=O)NC3=CC=CC=C3C(=O)N2C1</chem>
43	4-(4-Hydroxy-3,5-dimethoxyphenyl)-3-buten-2-one	<chem>CC(=O)C=CC1=CC(=C(C(=C1)OC)O)OC</chem>
44	Anthranilic acid	<chem>C1=CC=C(C(=C1)C(=O)O)N</chem>
45	Indirubin	<chem>C1=CC=C2C(=C1)C(=C(N2)O)C3=NC4=CC=CC=C4C3=O</chem>
46	Qingdaine	<chem>C1=CC=C2C(=C1)C(=O)/C(=C\3/C4=CC=CC=C4N5C3=NC6=CC=CC=C6C5=O)/N2</chem>
47	1,2-Dehydroreticuline	<chem>C[N+]1=C(C2=CC(=C(C=C2CC1)OC)O)CC3=CC(=C(C=C3)OC)O</chem>
48	Reticuline	<chem>CN1CCC2=CC(=C(C=C2[C@@H]1CC3=CC(=C(C=C3)OC)O)O)OC</chem>
49	Glaucine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=CC(=C(C=C4)OC)OC)OC)OC</chem>
50	N-methyllaurelliptine	<chem>[H][C@]12CC3=CC(O)=C(OC)C=C3C4=C1C(CCN2C)=CC(OC)=C4O</chem>
51	Laurelliptine	<chem>COC1=C(C2=C3C(CC4=CC(=C(C=C4)OC)O)NCCC3=C1)O</chem>
52	Hemiargyrine	<chem>c1(c(cc2c(c1)C[C@H]1N([C@H]2C)CCc2c1cc(c(c2)OC)O)OC)O</chem>
53	Corydine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)OC)O)OC</chem>
54	Isocorydine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)OC)OC</chem>
55	Antidesmone	<chem>CCCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C(=C(N2)C)OC</chem>
56	Hyeronine A	<chem>CCCCCCCC(=O)C1CCC2=C(C1)C(=C(NC2=O)C)OC</chem>

57	Hyeronine B	CCCCC(=O)C1CCC2=C(C1)C(=C(NC2=O)C)OC
58	14,15-Epoxy-norsecurinine	C1C[C@@H]2[C@]34CC(N2C1)[C@@H]5[C@H](C3=CC(=O)O4)O5
59	Norsecurinine	C1C[C@@H]2[C@]34C[C@H](N2C1)C=CC3=CC(=O)O4
60	Antidesmone	CCCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C(=C(N2)C)OC
61	Hyoscyamine	CN1[C@@H]2CC[C@H]1CC(C2)OC(=O)[C@H](CO)C3=CC=CC=C3
62	Calystegine B2	C1C[C@]2([C@H]([C@@H]([C@H]([C@@H]1N2)O)O)O)O
63	calystegine A3	C1C[C@]2([C@H]([C@@H](C[C@@H]1N2)O)O)O
64	Calystegine A5	C1C[C@]2(C[C@@H]([C@H]([C@@H]1N2)O)O)O
65	Calystegine B1	C1[C@H]2[C@@H](C[C@](N2)([C@H]([C@@H]1O)O)O)O
66	Calystegine B3	C1C[C@@]2([C@@H]([C@@H]([C@H]([C@H]1N2)O)O)O)O
67	(+)-Physoperuvine	CN1[C@@H]2CCC[C@]1(CC2)O
68	Physoperuvine	CN1C2CCCC1(CC2)O
69	Tropinone	CN1C2CCC1CC(=O)C2
70	Schizanthine A	CCCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C(=C(N2)C)OC
71	Schizanthine B	CC(=CC(=O)O[C@@H]1C[C@H]2C[C@H](C[C@@H]1N2C)OC(=O)/C=C(\C)/C(=O)O[C@@H]3C[C@@H]4C[C@H]([C@H](C3)N4C)OC(=O)C=C(C)C)C
72	Schizanthine C	C/C=C(\C)/C(=O)OC1CC2CC(CC1N2C)OC(=O)/C=C(\C)/C(=O)OC3CC4CCC(C3)N4C
73	Schizanthine D	C/C=C(/C)\C(=O)O[C@@H]1C[C@H]2C[C@H](C[C@@H]1N2C)OC(=O)/C=C(\C)/C(=O)O[C@@H]3C[C@H]4C[C@H]([C@H]3N4C)O
74	Grahamine	CCC(C)C(=O)OC1(C(C(=O)OC2CCN3C2C(=CC3)COC(=O)C1(C)O)C)C
75	Ecgonine	CN1[C@H]2CC[C@@H]1[C@H]([C@H](C2)O)C(=O)O
76	Cocaine	CN1[C@H]2CC[C@@H]1[C@H]([C@H](C2)OC(=O)C3=CC=CC=C3)C(=O)OC
77	Tropacocaine	CN1[C@@H]2CC[C@H]1CC(C2)OC(=O)C3=CC=CC=C3
78	α -Truxilline	CN1[C@H]2CC[C@@H]1[C@H]([C@H](C2)OC(=O)C3C(C(C3C4=CC=CC=C4)C(=O)O[C@H]5C[C@@H]6CC[C@H]([C@H]5C(=O)OC)N6C)C7=CC=CC=C7)C(=O)OC
79	Mooniine A	CCC(=O)O[C@@H]1C[C@H]2C[C@H](C[C@@H]1N2)OC(=O)C3C(C(C3C(=O)OC4C[C@H]5CC[C@@H](C4)N5C)C6=CC=CC=C6)C7=CC=CC=C7

80	Mooniine B	<chem>CC(=O)O[C@@H]1C[C@H]2C[C@H](C[C@@H]1N2C)OC(=O)C3C(C(C3C(=O)OC4C[C@H]5CC[C@@H](C4)N5C)C6=CC=CC=C6)C7=CC=CC=C7</chem>
81	Acetylknightino l	<chem>CC(=O)OC1CC2CCC(C1C(C3=CC=CC=C3)OC(=O)C)N2C</chem>
82	Strobamine	<chem>CN1[C@@H]2CC[C@H]1C3=C(C2)O[C@@H](CC3=O)C4=CC=CC=C4</chem>
83	chalcostrobamin e	<chem>CN1[C@@H]2CCC1CC(=O)/C2=C/C=C/C3=CC=CC=C3)\O</chem>
84	Brugine	<chem>CN1C2CCC1CC(C2)OC(=O)[C@@H]3CCSS3</chem>
85	Amuronine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1C[C@]34CCC(=O)C=C4)OC)OC</chem>
86	Antidesmone	<chem>CCCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C=C(N2)C)OC</chem>
88	Speranskilatine A	<chem>CC(=O)C[C@]1(C(=C(C(=O)N(C1=O)C)C2=C(C(=O)N(C2=O)C)OC)OC)O</chem>
89	Speranberculati ne A	<chem>CN1C(C(=C(C1=O)C2=C(C(=O)N(C2=O)C)OC)OC)O</chem>
90	Morphine	<chem>CN1CC[C@]23[C@@H]4[C@H]1CC5=C2C(=C(C=C5)O)O[C@H]3[C@H](C=C4)O</chem>
91	Aporphine	<chem>CN1CCC2=C3C1CC4=CC=CC=C4C3=CC=C2</chem>
92	Morphine	<chem>CN1CC[C@]23[C@@H]4[C@H]1CC5=C2C(=C(C=C5)O)O[C@H]3[C@H](C=C4)O</chem>
93	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
94	Aporphine	<chem>CN1CCC2=C3C1CC4=CC=CC=C4C3=CC=C2</chem>
95	(17RS)-17-(β- D- glucopyranosyl oxy)antidesmon e	<chem>COC1C(C)NC2C([C@@H](CCCCCCC(C)O[C@]34C[C@](O)(C3O)[C@H](O)C(CO)O4)CCC2=O)C1=O</chem>
96	(17RS)-8- deoxo-17-(β-D- glucopyranosyl oxy)antidesmon e	<chem>COC1C(C)NC2CCC[C@H](CCCCCCC(C)O[C@]34C[C@](O)(C3O)[C@H](O)C(CO)O4)C2C1=O</chem>
97	Aspertin-A	<chem>[H][C@@]1(C[C@@H](O)CCC)CC=C[C@@]([H])(C[C@@H](O)CC)N1</chem>
98	Aspertine C	<chem>CCC[C@@H](C[C@@H]1CCC[C@H](N1)C[C@H](CCC)O)O</chem>
99	Aspertin-B	<chem>[H]N1[C@]([H])(C[C@@]([H])(O)CC)CCC[C@@]1([H])C[C@@]([H])(O)CCCC</chem>

100	Aspertin-D	<chem>[H]N1[C@]([H])(C[C@@]([H])(O)CC)CCC[C@@]1([H])C[C@@]([H])(O)CCC</chem>
101	Lotthanongine	<chem>C1=CC(=CC=C1/C=C/C(=O)NCCC2=C(NC3=C2C=CC(=C3)O)[C@@H]4[C@@H]([C@H](OC5=CC(=CC(=C45)O)O)C6=CC=C(C=C6)O)O)O</chem>
102	17,18-Bis-nor-antidesmone	<chem>c1([nH]c(=O)c2c(c1OC)C[C@H](CC2)C(=O)CCCCC)C</chem>
103	18-Nor-antidesmone	<chem>CCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C(=C(N2)C)O</chem>
104	Antidesmone	<chem>CCCCCCCC[C@H]1CCC(=O)C2=C1C(=O)C(=C(N2)C)OC</chem>
105	8-Dihydroantidesmone	<chem>CCCCCCCC[C@H]1CC[C@H](C2=C1C(=O)C(=C(N2)C)OC)O</chem>
106	8-Deoxoantidesmone	<chem>CCCCCCCC[C@H]1CCCC2=C1C(=O)C(=C(N2)C)OC</chem>
107	N-1,N-2,N-3-triisopentenylguanidine	<chem>CC(=CCNC(=NCC=C(C)C)NCC=C(C)C)C</chem>
108	Julocrotine	<chem>CCC(C)C(=O)NC1CCC(=O)N(C1=O)CCC2=CC=CC=C2</chem>
109	Norsinoacutine	<chem>COC1=C(C2=C(C[C@@H]3C4=CC(=O)C(=C[C@@]42CCN3)OC)C=C1)O</chem>
110	Saludimerine A	<chem>CN1CC[C@@]23C=C(C(=O)C=C2[C@@H]1CC4=C3C(=C(C=C4C5=CC(=C(C6=C5C[C@H]7C8=CC(=O)C(=C[C@]86CCN7C)OC)O)OC)O)OC</chem>
111	3,4,8-trimethoxy-2-quinolone	<chem>[H]N1C(=O)C(OC)=C(OC)C2=C1C(OC)=CC=C2</chem>
112	2,3-methylenedioxy-4,7,8-trimethoxyquinoline	<chem>COC1=C(OC)C2=C(C=C1)C(OC)=C1OCOC1=N2</chem>
113	2,3-Methylenedioxy-4,8-dimethoxyquinoline	<chem>[H]C1=C(OC)C2=C(C=C1)C(OC)=C1OCOC1=N2</chem>

114	Muscicapine A	<chem>[H]N([H])C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(=CC1=O)C(C)C</chem>
115	muscicapine B	<chem>[H]N(CCC(C)C)C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(CC)=CC1=O</chem>
116	muscicapine C	<chem>[H]N(CC(C)CC)C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(=CC1=O)C(C)C</chem>
117	epiguaipyridine	<chem>CC(=C)[C@@H]2CC[C@@H](C)C=1C=CC(C)=NC=1C2</chem>
118	Anabasine	<chem>C1CCNC(C1)C2=CN=CC=C2</chem>
119	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
120	N1,N2,N3- Triisopentenyl guanidine (TIPG)	<chem>CC(=CCNC(=NCC=C(C)C)NCC=C(C)C)C</chem>
121	N1,N2- Diisopentenyl guanidine (DIPG)	<chem>CC(=CCNC(=NCC=C(C)C)N)C</chem>
122	Norsecurinine	<chem>C1C[C@@H]2[C@]34C[C@H](N2C1)C=CC3=CC(=O)O4</chem>
123	Securinine	<chem>C1CCN2[C@H](C1)[C@]34C[C@H]2C=CC3=CC(=O)O4</chem>
124	Phyllanthine	<chem>CO[C@@H]1CCN2[C@H](C1)C34C[C@H]2C=CC3=CC(=O)O4</chem>
125	Phyllochrysine	<chem>C1CCN2C(C1)C34CC2C=CC3=CC(=O)O4</chem>
126	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
127	glaucine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=CC(=C(C=C43)OC)OC)OC)OC</chem>
128	hemiargyrine	<chem>c1(c(cc2c(c1)C[C@H]1N([C@H]2C)CCc2c1cc(c(c2)OC)O)OC)O</chem>
129	salutaridine	<chem>CN1CC[C@]23C=C(C(=O)C=C2[C@H]1CC4=C3C(=C(C=C4)OC)O)OC</chem>
130	6-hydroxy-2- methyltetrahydr oharman	<chem>CC1N(C)CCC2=C1NC1=CC=C(C)C=C21</chem>
131	julocrotone	<chem>CC[C@H](C)C(=O)NC1CCC(=O)N(C[C@](C)(O)C2=CC=CC=C2)C1=O</chem>
132	Muscicapine A	<chem>[H]N([H])C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(=CC1=O)C(C)C</chem>
133	muscicapine B	<chem>[H]N(CCC(C)C)C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(CC)=CC1=O</chem>
134	muscicapine C	<chem>[H]N(CC(C)CC)C(\[H])=C1\C2=C(C(C)CC2)C(=O)C(=CC1=O)C(C)C</chem>
135	Virosecurinin	<chem>C1CCN2[C@@H](C1)[C@@]34C[C@@H]2C=CC3=CC(=O)O4</chem>

136	Viroallosecurinine	<chem>C1CCN2[C@H](C1)[C@@]34C[C@@H]2C=CC3=CC(=O)O4</chem>
137	14,15-Dihydronorsecurinine	<chem>C1C[C@@H]2[C@]34C[C@H](N2C1)CCC3=CC(=O)O4</chem>
138	ent-phyllanthidine	<chem>C1CCN2[C@H](C1)[C@@]34C[C@@H](O2)C=CC3=CC(=O)O4</chem>
139	Hordenine	<chem>CN(C)CCC1=CC=C(C=C1)O</chem>
140	Flueggeine A	<chem>C1C[C@@H]2[C@]34C[C@H](N2C1)C=C(C3=CC(=O)O4)[C@@H]5CC6=CC(=O)O[C@@]67C[C@@H]5N8[C@@H]7C CC8</chem>
141	Flueggeine B	<chem>C1C[C@@H]2[C@]34C[C@H](N2C1)[C@H]5[C@@H](C3=CC(=O)O4)[C@@H]6C/C(=C/C(=O)O)/[C@]7(C[C@@H]6[N +]58[C@@H]7CCC8)O</chem>
142	Mayteine	<chem>CC1C(C(=O)O[C@H]2[C@@H]([C@@H]([C@]3([C@@H]([C@@H]([C@@H]4[C@H]([C@@]3([C@@]2(C)O)O[C@]4(COC(=O)C5=C1N=CC=C5)C)OC(=O)C)OC(=O)C)OC(=O)C)COC(=O)C)OC(=O)C6=CC=CC=C6)OC(=O)C)C</chem>
143	Putterine A	<chem>[C@H]1([C@H]2[C@@]([C@@]34[C@]([C@@H]1OC(=O)C)([C@@H]([C@H]([C@@H]([C@@H]3OC(=O)C)[C@](O4)(C)COC(=O)c1c(nccc1)[C@H]([C@@H](C(=O)O2)C)C)OC(=O)c1cnccc1)OC(=O)C)COC(=O)C)(O)C)OC(=O)C</chem>
144	Putterine B	<chem>[C@@H]1([C@H]2[C@]([C@@]34[C@]([C@H]1OC(=O)C)([C@@H]([C@@H]([C@@H]([C@@H]3OC(=O)C)[C@](O4)(C) COC(=O)c1c(nccc1)[C@H]([C@@H](C(=O)O2)C)C)OC(=O)c1cnccc1)OC(=O)C)COC(=O)C)(O)C)OC(=O)C(C)C</chem>
145	Bisbenzylisoquinoline ether	<chem>C1=CC=C(C=C1)CC2=C3C=CC=C4C3=C(O4)C(=N2)CC5=CC=CC=C5</chem>
146	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
147	N1,N2-Diisopentenylguanidine (DIPG)	<chem>CC(=CCNC(=NCC=C(C)C)N)C</chem>
148	N1,N2,N3-Triisopentenylguanidine (TIPG)	<chem>CC(=CCNC(=NCC=C(C)C)NCC=C(C)C)C</chem>
149	Magnoflorine	<chem>C[N+]1(CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)O)OC)C</chem>
150	Trigonostemonine A	<chem>[H]C1=CC2=C(C=C1)C1=C(N2)C(=O)N(CCC(=O)C2=CC=C(OC)C=C2N)CC1</chem>
151	Trigonostemonine B	<chem>[H]C1=CC=C(C(=O)CCN2CCC3=C(NC4=C3C=CC(OC)=C4)C2=O)C(N)=C1</chem>

152	Trigonostemonine C	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C4=CC=NC5=CC=CC=C45</chem>
153	Trigonostemonine D	<chem>COC1=CC2=NC=CC(=C2C=C1)C3=NC=CC4=C3NC5=CC=CC=C45</chem>
154	Trigonostemonine E	<chem>COC1=CC2=NC=CC(=C2C=C1)C3=NCCC4=C3NC5=CC=CC=C45</chem>
155	Trigonostemonine F	<chem>COC1=CC2=C(C=C1)C1=C(N2)C2N(CCC3=C2NC2=C3C=CC=C2)CC1</chem>
156	2-Episecurinol A	<chem>C1CCN2[C@@H](C1)[C@@]34C[C@H]([C@@H]2CC3=CC(=O)O4)O</chem>
157	Allosecurinine	<chem>C1CCN2[C@@H](C1)[C@]34CC2C=CC3=CC(=O)O4</chem>
158	Virosecurinine	<chem>C1CCN2[C@@H](C1)[C@@]34C[C@H]2C=CC3=CC(=O)O4</chem>
159	Securinine	<chem>CO[C@H]1CCN2[C@@H](C1)[C@]34C[C@H]2C=CC3=CC(=O)O4</chem>
160	15 α -methoxy-14,15-dihydrophyllorhynchysine	<chem>[H][C@@]12CCCCN1[C@H]1C[C@@]22OC(=O)C=C2C[C@@H]1OC</chem>
161	ent-phyllanthidine	<chem>C1CCN2[C@H](C1)[C@@]34C[C@H](O2)C=CC3=CC(=O)O4</chem>
162	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
163	Houttuynamide A	<chem>C1=CC(=CC=C1CCNC(=O)C2=CC(=C(C=C2)O)O)O</chem>
164	Secu'amamine B	<chem>CO[C@H]1CCN2[C@@H](C1)[C@]34C[C@H]2[C@@H](CC3=CC(=O)O4)OC</chem>
165	Secu'amamine C	<chem>CO[C@@H]1CC2=CC(=O)O[C@@]23C[C@@H]1N4[C@@H]3CCCC4</chem>
166	fluggeainol	<chem>[H][C@]12CCCCN1[C@H]1C[C@@]22OC(=O)C=C2C[C@]1([H])O</chem>
167	4-ALPHA-HYDROXYALLOSECURININE	<chem>[H][C@@]12C[C@H](CCN1[C@H]1C[C@@]22OC(=O)C=C2C=C1)OC</chem>
168	4- α -Hydroxy-15a-methoxy-14,15-	<chem>[H]C12C[C@@H](O)CCN1[C@H]1C[C@@]22OC(=O)C=C2C=C1OC</chem>

	dihydroallosecurinine	
169	4 α -Methoxy-15 α -methoxy-14,15-dihydrosecurinine	<chem>[H]C12C[C@H](CCN1[C@H]1C[C@@]22OC(=O)C=C2C=C1OC)OC</chem>
170	14 β -Hydroxy-13,14-dihydronorsecurinine	<chem>[H][C@]12CCCCN1[C@H]1C[C@@]22OC(=O)C=C2C[C@@]1([H])O</chem>
171	14 β -Methoxy-13,14-dihydronorsecurinine	<chem>[H][C@]12CCCCN1[C@H]1C[C@@]22OC(=O)C=C2C[C@@]1([H])OC</chem>
172	14 α -Methoxy-13,14-dihydronorsecurinine	<chem>[H][C@]12CCCCN1[C@H]1C[C@@]22OC(=O)C=C2C[C@]1([H])OC</chem>
173	1,5-Diphenyl-3-styryl-2-pyrazoline	<chem>C1C(N(N=C1/C=C/C2=CC=CC=C2)C3=CC=CC=C3)C4=CC=CC=C4</chem>
174	Uracil	<chem>C1=CNC(=O)NC1=O</chem>
175	Uridine	<chem>C1=CN(C(=O)NC1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O</chem>
176	2-(5-methyl-1H-indol-3-yl)ethanol	<chem>CC1=CC2=C(C=C1)NC=C2CCO</chem>
177	julocrotine	<chem>CCC(C)C(=O)NC1CCC(=O)N(C1=O)CCC2=CC=CC=C2</chem>
178	Tropane	<chem>CN1[C@@H]2CCC[C@H]1CC2</chem>
179	Piperidine	<chem>C1CCNCC1</chem>
180	Piperidine-2,6-dione	<chem>C1CC(=O)NC(=O)C1</chem>
181	pyridinone	<chem>C1=CC(=O)NC=C1</chem>

182	Hermidin	<chem>CN1C(=O)CC(=C(C1=O)O)OC</chem>
183	Pseudoakummi gine	<chem>C/C=C/I\CN2CCC34C5=CC=CC=C5N(C36C2CC1C4(CO6)C(=O)OC)C</chem>
184	Berberine	<chem>COC1=C(C2=C[N+]3=C(C=C2C=C1)C4=CC5=C(C=C4CC3)OCO5)OC</chem>
185	Palmatine	<chem>COC1=C(C2=C[N+]3=C(C=C2C=C1)C4=CC(=C(C=C4CC3)OC)OC)OC</chem>
186	Jatrorrhizine	<chem>COC1=C(C2=C[N+]3=C(C=C2C=C1)C4=CC(=C(C=C4CC3)O)OC)OC</chem>
187	Epiberberine	<chem>COC1=C(C=C2C(=C1)CC[N+]3=C2C=C4C=CC5=C(C4=C3)OCO5)OC</chem>
188	Coptisine	<chem>C1C[N+]2=C(C=C3C=CC4=C(C3=C2)OCO4)C5=CC6=C(C=C51)OCO6</chem>
189	Groenlandicine	<chem>COC1=C(C=C2CC[N+]3=C(C2=C1)C=C4C=CC5=C(C4=C3)OCO5)O</chem>
190	Magnoflorine	<chem>C[N+]1(CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)O)OC)C</chem>
191	Lupanine	<chem>C1CCN2C[C@@H]3C[C@H]([C@@H]2C1)CN4[C@@H]3CCCC4=O</chem>
192	13- Hydroxylupanin e	<chem>C1CC2C3CC(CN2C(=O)C1)C4CC(CCN4C3)O</chem>
193	17-Oxolupanine	<chem>C1CCN2[C@@H](C1)[C@H]3C[C@@H](C2=O)[C@H]4CCCC(=O)N4C3</chem>
194	TETRAMETH YLPYRAZINE	<chem>CC1=C(N=C(C(=N1)C)C)C</chem>
195	Rhetsinine	<chem>CNC1=CC=CC=C1C(=O)N2CCC3=C(C2=O)NC4=CC=CC=C34</chem>
196	Securinine	<chem>C1CCN2[C@H](C1)[C@]34C[C@H]2C=CC3=CC(=O)O4</chem>
197	Norsecurinine	<chem>C1C[C@@H]2[C@]34C[C@H](N2C1)C=CC3=CC(=O)O4</chem>
198	Viroallosecurini ne	<chem>C1CCN2[C@H](C1)[C@@]34C[C@@H]2C=CC3=CC(=O)O4</chem>
199	Isosalutaridine	<chem>CN1CC[C@@]23C=C(C(=O)C=C2[C@@H]1CC4=CC(=C(C=C34)OC)O)OC</chem>
200	Flavinantine	<chem>CN1CC[C@]23C=C(C(=O)C=C2[C@H]1CC4=CC(=C(C=C34)O)OC)OC</chem>
201	o- Methylflavinant ine	<chem>CN1CC[C@]23C=C(C(=O)C=C2[C@H]1CC4=CC(=C(C=C34)OC)OC)OC</chem>
202	Isocorydine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)OC)OC</chem>
203	3-hydroxy-2,4- dimethoxy-10-	<chem>COC1=CC2=C(N(C)C3=C(C=CC=C3)C2=O)C(OC)=C1O</chem>

	methyl-9-acridanone	
204	Acridanone	<chem>C1=CC=C2C(=C1)C(=O)C3=CC=CC=C3N2</chem>
205	Crotsparsidine	<chem>COC1=C(C=C2C3=C1CCN[C@H]3CC24C=CC(=O)C=C4)O</chem>
206	Choline	<chem>C[N+](C)(C)CCO</chem>
207	Trigonelline	<chem>C[N+]1=CC=CC(=C1)C(=O)[O-]</chem>
208	Angustifoline	<chem>C=CC[C@H]1[C@H]2C[C@@H](CN1)[C@H]3CCCC(=O)N3C2</chem>
209	Sparteine	<chem>C1CCN2C[C@@H]3C[C@H]([C@H]2C1)CN4[C@H]3CCCC4</chem>
210	9-Octadecenamide	<chem>CCCCCCCCC=CCCCCCCC(=O)N</chem>
211	Dihydro-oxodemethoxyhemanthamine	<chem>C1CCC23C(C1)N(CC4=CC5=C(C=C42)OCO5)CC3=O</chem>
212	Augustamine	<chem>CN1CCC23C1CCC4C2OC(O4)C5=CC6=C(C=C35)OCO6</chem>
213	Lupanine	<chem>C1CCN2C[C@@H]3C[C@H]([C@@H]2C1)CN4[C@@H]3CCCC4=O</chem>
214	Oxoasoanine	<chem>COC1=C(C=C2C(=C1)C3=CC=CC4=C3N(C2=O)CC4)OC</chem>
215	Cinchonidine	<chem>C=C[C@H]1CN2CC[C@H]1C[C@H]2[C@@H](C3=CC=NC4=CC=CC=C34)O</chem>
216	Cinchonine	<chem>C=C[C@H]1CN2CC[C@H]1C[C@@H]2[C@H](C3=CC=NC4=CC=CC=C34)O</chem>
217	Buphandrin	<chem>CO[C@@H]1C[C@@H]2[C@@]3(CCN2CC4=C(C5=C(C=C43)OCO5)OC)C=C1</chem>
218	Indicine N-oxide	<chem>C[C@@H]([C@](C(C)C)(C(=O)OCC1=CC[N+]2([C@H]1[C@@H](CC2)O)[O-])O)O</chem>
219	Powellin	<chem>COC1=C2CN3CC[C@@]4([C@H]3C[C@H](C=C4)O)C2=CC5=C1OCO5</chem>
220	Undulatine	<chem>CO[C@@H]1C[C@@H]2[C@@]3(CCN2CC4=C(C5=C(C=C43)OCO5)OC)[C@@H]6[C@H]1O6</chem>
221	Ambelline	<chem>CO[C@@H]1C[C@@H]2[C@@]3(C=C1)[C@H](CN2CC4=C(C5=C(C=C34)OCO5)OC)O</chem>
222	6-Hydroxybuphanidrine	<chem>COc1c2c(cc3c1OCO3)C45CCN(C4CC(C=C5)OC)C2O</chem>
223	Acronine	<chem>CC1(C=CC2=C3C(=C(C=C2O1)OC)C(=O)C4=CC=CC=C4N3C)C</chem>
224	Monocrotaline	<chem>C[C@H]1C(=O)O[C@@H]2CCN3[C@@H]2C(=CC3)COC(=O)[C@]([C@]1(C)O)(C)O</chem>

225	6-Hydroxypowelline	<chem>COc1c2c(cc3c1OCO3)C45CCN(C4CC(C=C5)O)C2O</chem>
226	Nitidine	<chem>C[N+]1=CC2=CC(=C(C=C2C3=C1C4=CC5=C(C=C4C=C3)OCO5)OC)OC</chem>
227	Crinamidine	<chem>COC1=C2CN3CC[C@@]4([C@H]3C[C@H]([C@H]5[C@@H]4O5)O)C2=CC6=C1OCO6</chem>
228	Cavinine	<chem>CO[C@@H]1C[C@@H]2[C@@]3([C@@H](CN2CC4=C(C5=C(C=C43)OCO5)OC)O)[C@@H]6[C@H]1O6</chem>
229	6-Hydroxyundulatine	<chem>CO[C@@H]1C[C@@H]2[C@@]3(CCN2C(C4=C(C5=C(C=C43)OCO5)OC)O)[C@@H]6[C@H]1O6</chem>
230	Epoxy-3,7-Dimethoxycrinane 11-one	<chem>COc1c2c(cc3c1OCO3)C45C(CC(C6C4O6)OC)N(C2)CC5=O</chem>
231	Echitamidin	<chem>C[C@@H]([C@@H]1CN2CCC34[C@@H]2C[C@@H]1C(=C3NC5=CC=CC=C45)C(=O)OC)O</chem>
232	Voacangine	<chem>CC[C@H]1C[C@H]2C[C@@]3([C@H]1N(C2)CCC4=C3NC5=C4C=C(C=C5)OC)C(=O)OC</chem>
233	Mitraphylline	<chem>C[C@H]1[C@H]2CN3CC[C@]4([C@@H]3C[C@@H]2C(=CO1)C(=O)OC)C5=CC=CC=C5NC4=O</chem>
234	Camptothecin	<chem>CC[C@@]1(C2=C(COC1=O)C(=O)N3CC4=CC5=CC=CC=C5N=C4C3=C2)O</chem>
235	Echitamine	<chem>C/C=C\1/C[N@@+]2(CC[C@@]34[C@@]2([C@H](C[C@@H]1C3(CO)C(=O)OC)O)NC5=CC=CC=C45)C</chem>
236	Colchicine	<chem>CC(=O)N[C@H]1CCC2=CC(=C(C=C2C3=CC=C(C(=O)C=C13)OC)OC)OC</chem>
237	Emetine	<chem>CC[C@H]1CN2CCC3=CC(=C(C=C3[C@@H]2C[C@@H]1C[C@@H]4C5=CC(=C(C=C5CCN4)OC)OC)OC)OC</chem>
238	Tetrandrine	<chem>CN1CCC2=CC(=C3C=C2[C@@H]1CC4=CC=C(C=C4)OC5=C(C=CC(=C5)C[C@H]6C7=C(O3)C(=C(C=C7CCN6C)OC)OC)OC)OC</chem>
239	Thalicarpin	<chem>CN1CCC2=CC(=C(C3=C2C1CC4=CC(=C(C=C43)OC)OC)OC5=CC(=C(C=C5CC6C7=CC(=C(C=C7CCN6C)OC)OC)OC)OC)OC)OC</chem>
240	Paclitaxel	<chem>CC1=C2[C@H](C(=O)[C@@]3([C@H](C[C@@H]4[C@]([C@H]3[C@@H]([C@@](C2(C)C)(C[C@@H]1OC(=O)[C@@H]([C@H](C5=CC=CC=C5)NC(=O)C6=CC=CC=C6)O)O)OC(=O)C7=CC=CC=C7)(CO4)OC(=O)C)O)C)OC(=O)C</chem>
241	Corydine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)OC)O)OC</chem>
242	Salutaridine	<chem>CN1CC[C@]23C=C(C(=O)C=C2[C@H]1CC4=C3C(=C(C=C4)OC)O)OC</chem>
243	Aporphine	<chem>CN1CCC2=C3C1CC4=CC=CC=C4C3=CC=C2</chem>
244	Phenazine	<chem>C1=CC=C2C(=C1)N=C3C=CC=CC3=N2</chem>
245	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
246	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>

247	julocrotine	<chem>CCC(C)C(=O)NC1CCC(=O)N(C1=O)CCC2=CC=CC=C2</chem>
248	crotonimide B	<chem>CC(C)C(=O)N[C@H]1CCC(=O)N(CCC2=CC=CC=C2)C1=O</chem>
249	crotonimide A	<chem>CCC(=O)N[C@H]1CCC(=O)N(CCC2=CC=CC=C2)C1=O</chem>
250	N-[2,6-dioxo-1-(2-phenylethyl)-3-piperidiny]-acetamide	<chem>CC(=O)N(C1CCN(CCC2=CC=CC=C2)CC1)C1=CC=CC=C1</chem>
251	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
252	Curcacycline A	<chem>C[C@H]([C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N1)CC(C)C)CC(C)C)CC(C)C)CC(C)C)C(C)C)O</chem>
253	Curcacycline B	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N1)CC(C)C)CC(C)C)[C@@H](C)CC)CO)CC(C)C</chem>
254	Jatrophidin I	<chem>CC(C)C[C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N2CCC[C@H]2C(=O)NCC(=O)N1)CC3=CNC4=CC=CC=C43)CC(C)C)CC(=O)N)CC(C)C</chem>
255	Chevalierin A	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N1)C)CC(C)C)[C@@H](C)CC)[C@@H](C)CC)CCSC</chem>
256	Chevalierin B	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N1)C)CC(C)C)[C@@H](C)CC)[C@@H](C)CC)CCS(=O)C</chem>
257	Chevalierin C	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N1)CC(=O)O)CC2=CC=CC=C2)[C@@H](C)CC)[C@@H](C)O)CC3=CC=C(C=C3)O)C)CC4=CC=CC=C4</chem>
258	Cyclogossine B	<chem>CCC(C)C1C(=O)NC(C(=O)NCC(=O)NCC(=O)NC(C(=O)NC(C(=O)NC(C(=O)NC(C(=O)N1)C)C)CC(C)C)CC2=CNC3=CC=CC=C32)CC(C)C</chem>
259	Integerrimide A	<chem>C[C@H]([C@H]1C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N1)CC(C)C)CC(C)C)CC(C)C)CC3=CNC4=CC=CC=C43)O</chem>
260	Integerrimide B	<chem>C[C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)N1)C3=CNC4=CC=CC=C43)CO)C(C)C)CC(C)C)CC(C)C</chem>
261	Mahafacyclin A	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N1)[C@@H](C)O)CC2=CC=CC=C2)C(C)C)CC(C)C</chem>
262	Mahafacyclin B	<chem>C[C@H]([C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N1)CC2=CC=C(C=C2)CC3=CC=CC=C3)CC4=CC=CC=C4)CC5=CC=CC=C5)O</chem>
263	Podacycline A	<chem>C[C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N1)CC(C)C)CC(C)C)CC2=CNC3=CC=CC=C32)C(C)C</chem>

264	Podacycline B	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)N1)[C@@H](C)O)C)CC2=CC=CC=C2)CC3=CC=CC=C3</chem>
265	Pohlianin A	<chem>CC(C)C[C@H]1C(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N1)C3=CC=C(C=C3)O)CC(C)C)CC(C)C)C(C)C</chem>
266	Pohlianin B	<chem>CC(C)C[C@H]1C(=O)N[C@H](C(=O)N[C@H](C(=O)N[C@H](C(=O)N2CCC[C@H]2C(=O)N[C@H](C(=O)NCC(=O)N1)C(C)C)CC3=CC=C(C=C3)O)CC(C)C)CC(C)C</chem>
267	Pohlianin C	<chem>CC[C@H](C)[C@H]1C(=O)N[C@H](C(=O)NCC(=O)N[C@H](C(=O)NCC(=O)NCC(=O)N[C@H](C(=O)N[C@H](C(=O)N1)[C@@H](C)CC)[C@@H](C)O)CC2=CC=CC=C2)CC3=CC=CC=C3</chem>
268	Norocobitrine	<chem>[H]N1CC[C@]23C=C(OC)C(=O)C[C@]2([H])[C@H]1CC1=CC=C(OC)C(O)=C31</chem>
269	Nildahilline	<chem>[H]N1CC[C@]23C=C(OC)C(=O)C[C@@]2([H])[C@H]1CC1=CC=C(OC)C(O)=C31</chem>
270	crotonimide C	<chem>[H]N(C(=O)C1=CC=CC=C1)[C@@]1([H])CCC(=O)N(CCC2=CC=CC=C2)C1=O</chem>
271	Julocrotine	<chem>CCC(C)C(=O)NC1CCC(=O)N(C1=O)CCC2=CC=CC=C2</chem>
272	Colchicine	<chem>CC(=O)N[C@H]1CCC2=CC(=C(C=C2C3=CC=C(C(=O)C=C13)OC)OC)OC</chem>
273	5-Hydroxypyrrolidin-2-one	<chem>C1CC(=O)NC1O</chem>
274	Diamide	<chem>CN(C)C(=O)/N=N/C(=O)N(C)C</chem>
275	Uracil	<chem>C1=CNC(=O)NC1=O</chem>
276	Hermidin	<chem>CN1C(=O)CC(=C(C1=O)O)OC</chem>
277	Cyanohermidin	<chem>CN1C(=O)C=C(C(=O)C1=O)OC</chem>
278	hermidin quinone	<chem>COC1=CC(=O)N(C)C(=O)C1=O</chem>
279	TAS2C	<chem>COC1=C2OC(=O)C3=C4C(OC(=O)C(C(CCC(=O)NCCN5CCCCC5)=C1)=C24)=C(OC)C=C3CCC(=O)NCCN1CCCCC1</chem>
280	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
281	7-deoxy-homoDMDP	<chem>C[C@@H]([C@@H]1[C@H]([C@@H]([C@H](N1)CO)O)O)O</chem>
282	alpha-Homonojirimycin	<chem>C([C@@H]1[C@H](C([C@H]([C@H](N1)CO)O)O)O)O</chem>
283	N-Methyl-alpha-	<chem>CN1[C@@H]([C@H](C([C@H]([C@H]1CO)O)O)O)CO</chem>

	homonojirimycin	
284	alpha-Homomannojirimycin	<chem>C([C@@H]1[C@H](C([C@@H]([C@H](N1)CO)O)O)O)O</chem>
285	beta-homomannojirimycin	<chem>C([C@@H]1[C@H](C([C@@H]([C@@H](N1)CO)O)O)O)O</chem>
286	alfa-7-desoxy-homonojirimicina	<chem>[H]OC[C@H]1N([H])[C@H](C)[C@H](O[H])C[C@@H]1O[H]</chem>
287	7-O-beta-D-Glucopiranosil-alfa-homonojirimicina	<chem>C(C1C(C(C(C(N1)COC2C(C(C(C(O2))))O)O)O)O)O)O</chem>
288	Sparsifloamide	<chem>CCCCCCCC\C=C/[C@@H](O)C(=O)NC(CO)[C@H](O)[C@H](O)\C=C\CC\C=C/CCCCCCC</chem>
289	Jatrofham	<chem>CC1=C[C@H](NC1=O)O</chem>
290	guanidine	<chem>CC(C)=CCNC1=NO[C@H](CN1)C(C)=C</chem>
291	alchornedine	<chem>CC(C)=CCNC1=NOC(CN1)C(C)=C</chem>
292	pteroginine	<chem>CC(C)=CCNC(=N)NCC=C(C)C</chem>
293	pteroginidine	<chem>CC(C)=CCN(CC=C(C)C)C(N)=N</chem>
294	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
295	7-deoxy-homoDMDP	<chem>C[C@@H](O)[C@H]1N[C@H](CO)[C@@H](O)[C@@H]1O</chem>
296	6-deoxy-homoDMDP	<chem>OCC[C@H]1N[C@H](CO)[C@@H](O)[C@@H]1O</chem>
297	2,5-imino-2,5,6-trideoxy-D-gulo-heptitol	<chem>OCC[C@H]1N[C@@H](CO)[C@H](O)[C@H]1O</chem>
298	4-deoxy-homoDMDP4-	<chem>OC[C@@H](O)[C@@H]1C[C@H](O)[C@@H](CO)N1</chem>

	deoxy-homoDMDP	
300	2,5-imino-2,4,5-trideoxy-D-gulo-heptitol	<chem>OC[C@@H](O)[C@@H]1C[C@@H](O)[C@H](CO)N1</chem>
301	2,5-imino-2,4,5,6-tetradideoxy-D-gulo-heptitol	<chem>OCC[C@@H]1C[C@@H](O)[C@H](CO)N1</chem>
302	6-C-butyl-DMDP	<chem>CCCC[C@H](O)[C@H]1N[C@H](CO)[C@@H](O)[C@@H]1O</chem>
303	6-C-butyl-4-deoxyDMDP	<chem>[H]C1[C@H](O)[C@@H](CO)N[C@@H]1[C@@H](O)CCCC</chem>
304	6-C-(8-hydroxyoctyl)-DMDP	<chem>OCC[C@@H](O)[C@H]1N[C@H](CO)[C@@H](O)[C@@H]1O</chem>
305	6-C-(8-hydroxyoctyl)-2,5-dideoxy-2,5-imino-D-galactitol	<chem>OCC[C@@H](O)[C@H]1N[C@@H](CO)[C@@H](O)[C@H]1O</chem>
306	6-Hydroxy-1-methyl-2-dimethyl-3,4-tetrahydro-beta-carboline	<chem>CC1C2=C(CC[N+]1(C)C)C1=CC(O)=CC=C1N2</chem>
307	Magnoflorine	<chem>C[N+]1(CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)OC)C</chem>
308	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
309	moschamine	<chem>COC1=C(C=CC(=C1)/C=C/C(=O)NCCC2=CNC3=C2C=C(C=C3)O)O</chem>
310	Trigonostemonine A	<chem>[H]C1=CC2=C(C=C1)C1=C(N2)C(=O)N(CCC(=O)C2=CC=C(OC)C=C2N)CC1</chem>
311	Trigonostemonine B	<chem>[H]C1=CC=C(C(=O)CCN2CCC3=C(NC4=C3C=CC(OC)=C4)C2=O)C(N)=C1</chem>

312	Trigonostemone C	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C4=CC=NC5=CC=CC=C45</chem>
313	Trigonostemone D	<chem>COC1=CC2=NC=CC(=C2C=C1)C3=NC=CC4=C3NC5=CC=CC=C45</chem>
314	Trigonostemone E	<chem>COC1=CC2=NC=CC(=C2C=C1)C3=NCCC4=C3NC5=CC=CC=C45</chem>
315	Trigonostemone F	<chem>COC1=CC2=C(C=C1)C1=C(N2)C2N(CCC3=C2NC2=C3C=CC=C2)CC1</chem>
316	11-Methoxyrutecarpine	<chem>COC1=CC2=C(C=C1)C3=C(N2)C4=NC5=CC=CC=C5C(=O)N4CC3</chem>
317	2-Methoxyruteacarpine	<chem>COC1=CC2=C(C=C1)C(=O)N3CCC4=C(C3=N2)NC5=CC=CC=C45</chem>
318	Trigonostemine A	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)C4=CNC5=CC=CC=C54</chem>
319	Trigonostemine B	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)C4=CNC5=C4C=CC(=C5)OC</chem>
320	Trigonostemine C	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)C4=C(NC5=CC=CC=C54)O</chem>
321	Trigonostemine D	<chem>COC1=CC2=C(C=C1)C(=C(N2)O)C(=O)C3=NC=CC4=C3NC5=CC=CC=C45</chem>
322	Trigonostemine E	<chem>COC1=CC2=C(C=C1)C3=CC(=O)NC(=C3N2)C4=CC=NC5=CC=CC=C45</chem>
323	Trigonostemine F	<chem>COC1=CC2=NC=CC(=C2C=C1)C3=C4C(=CC(=O)N3)C5=CC=CC=C5N4</chem>
324	arenarine c	<chem>CC(=O)C1=NC=CC2=C1NC3=C2C=CC(=C3)OC</chem>
325	Harmic Acid Methyl Ester	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)OC</chem>
326	1,2,3,4-tetrahydro-oxo-β-carboline	<chem>[H]C1=NC=CC2=C1NC1=CC(OC)=CC=C21</chem>
327	NORHARMIN E	<chem>COC1=CC2=C(C=C1)C3=C(N2)C=NC=C3</chem>

328	Harmine	<chem>CC1=NC=CC2=C1NC3=C2C=CC(=C3)OC</chem>
329	Trigonoine B	<chem>COC1=CC2=NC(=C3C(=C2C=C1)C=CN3)N4CCC(=O)C5=CC=CC=C54</chem>
330	Trigonoine A	<chem>COC1=CC2=C(C=C1)C3=C4N2[C@@H]5C[C@H](C4=NC=C3)C6=CC=CC=C6N5</chem>
331	trifiline A	<chem>[H]N1C2=CC(OC)=CC=C2C2=C1C1N(CCC3=C1NC1=C3C=CC(OC)=C1)CC2</chem>
332	trifiline B	<chem>COC1=CC2=C(C=C1)C1=C(N2)C2N(CC1)CCC1=C2N(CO)C2=CC(OC)=CC=C12</chem>
333	trifiline C	<chem>COC1=CC2=C(C=C1)C1CC(=O)N3CCC4=C(NC5=CC(OC)=CC=C45)C3=C1N2</chem>
334	Trigonoliimine A	<chem>COC1=CC2=C(C=C1)[C@@]34CCN=C3C5=C(CCN4C=N2)C6=CC=CC=C6N5</chem>
335	Trigonoliimine B	<chem>COC1=CC2=C(C=C1)C3=C(N2)C4=NCC[C@]45C6=CC=CC=C6N=CN5CC3</chem>
336	Trigonoliimine C	<chem>COC1=CC2=C(C=C1)C3=NCCC4=C([C@]3(N2)CCNC=O)NC5=CC=CC=C45</chem>
337	Trigonostemon F	<chem>CC(=O)/C(=C/1\C2=C(C=C(C=C2)OC)NC1=O)/C3=CNC4=CC=CC=C43</chem>
338	Trigolutesin A	<chem>C1C(=O)NC2=C([C@@]13[C@@H]([C@H](OC3=O)CO)C4=CNC5=CC=CC=C54)C=CC(=C2)O</chem>
339	Trigolutesin B	<chem>COCN1C=C(C2=CC=CC=C21)[C@@H]3[C@H](OC(=O)[C@@]34CC(=O)NC5=C4C=CC(=C5)O)CO</chem>
340	Trigolute A	<chem>C1C(=O)O[C@@H]([C@H]([C@]12C3=C(C=C(C=C3)O)N(C2=O)CO)C4=CNC5=CC=CC=C54)CO</chem>
341	Trigolute B	<chem>C1C(=O)O[C@@H]([C@H]([C@]12C3=C(C=C(C=C3)O)NC2=O)C4=CNC5=CC=CC=C54)CO</chem>
342	Trigolute C	<chem>COCN1C=C(C2=CC=CC=C21)[C@@H]3[C@H](OC(=O)[C@@]34C5=C(C=C(C=C5)O)NC4=O)CO</chem>
343	Trigolute D	<chem>CO[C@H]1C[C@]2([C@@H]([C@H](O1)CO)C3=CNC4=CC=CC=C43)C5=C(C=C(C=C5)O)NC2=O</chem>
344	Lotthanongine	<chem>C1=CC(=CC=C1/C=C/C(=O)NCCC2=C(NC3=C2C=CC(=C3)O)[C@@H]4[C@@H]([C@H](OC5=CC(=CC(=C45)O)O)C6=CC=C(C=C6)O)O)O</chem>
345	Bukittinggine	<chem>CC1CN2[C@H]3[C@H]4CCC35CCC[C@H]5[C@@]67C2[C@@H]1CC[C@@]46COC(=O)CC7</chem>
346	Vallesamine	<chem>C/C=C/1\CN2CC[C@@H]1[C@@](C3=C(C2)C4=CC=CC=C4N3)(CO)C(=O)OC</chem>
347	Ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
348	N-trans-Feruloyl-3,5-dihydroxyindolin-2-one	<chem>COC1=CC(\C=C\C(=O)NCCC2(C)C(=O)NC3=C2C=C(O)C=C3)=CC=C1O</chem>

349	N-trans-p-Coumaroyl-tryptamine	<chem>[H]C1=CC2=C(NC=C2CCNC(=O)\C=C\C2=CC=C(O)C([H])=C2)C=C1</chem>
350	N-trans-p-Coumaroyl-5-hydroxytryptamine	<chem>[H]C1=CC(\C=C\C(=O)NCCC2=CNC3=C2C=C(O)C=C3)=CC=C1O</chem>
351	N-trans-4-Methoxycinnamoyl-5-hydroxytryptamine	<chem>[H]C1=CC(\C=C\C(=O)NCCC2=CNC3=C2C=C(O)C=C3)=CC=C1OC</chem>
352	N-trans-Feruloyl-5-hydroxytryptamine	<chem>COC1=CC(\C=C\C(=O)NCCC2=CNC3=C2C=C(O)C=C3)=CC=C1O</chem>
353	1-(7-methoxyquinolinyl-4'-yl)-3,4-dihydro-β-carboline	<chem>COC1=CC=C2C(NC3=C2CCN=C3C2=C3C=CC=CC3=NC=C2)=C1</chem>
354	Corydine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)OC)O)OC</chem>
355	Norisoboldine	<chem>COC1=C(C2=C3[C@H](CC4=CC(=C(C=C4)OC)O)NCCC3=C1)O</chem>
356	Tembetarine	<chem>C[N+]1(CCC2=CC(=C(C=C2[C@@H]1CC3=CC(=C(C=C3)OC)O)O)OC)C</chem>
357	Magnoflorine	<chem>C[N+]1(CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)O)OC)C</chem>
358	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
359	Benzo[h]quinoline	<chem>C1=CC=C2C(=C1)C=CC3=C2N=CC=C3</chem>
360	Indole	<chem>C1=CC=C2C(=C1)C=CN2</chem>
361	Yohimbine	<chem>COC(=O)[C@H]1[C@H](CC[C@@H]2[C@@H]1C[C@H]3C4=C(CCN3C2)C5=CC=CC=C5N4)O</chem>
362	N1,N2-diisopentenylguanidine	<chem>CC(=C)CCN\C(N)=N\CCC(C)=C</chem>

363	N1,N2,N3-triisopentenyl guanidine	<chem>CC(=C)CCNC(NCCC(C)=C)=NCCC(C)=C</chem>
364	Indomethacin	<chem>CC1=C(C2=C(N1C(=O)C3=CC=C(C=C3)Cl)C=CC(=C2)OC)CC(=O)O</chem>
365	Guanidine, 1,2,3-tris(3-methyl-2-butenyl)- (8CI) or Alchorneine	<chem>CC(C)=CCNC(NCC=C(C)C)=NCC=C(C)C</chem>
366	Vincristine	<chem>CC[C@@]1(C[C@@H]2[C@@](C3=C(CCN(C2)C1)C4=CC=CC=C4N3)(C5=C(C=C6C(=C5)[C@]78CCN9[C@H]7[C@@](C=CC9)([C@H]([C@@]([C@@H]8N6C=O)(C(=O)OC)O)OC(=O)C)CC)OC)C(=O)OC)O</chem>
367	Vinblastine	<chem>CC[C@@]1(C[C@@H]2[C@@](C3=C(CCN(C2)C1)C4=CC=CC=C4N3)(C5=C(C=C6C(=C5)[C@]78CCN9[C@H]7[C@@](C=CC9)([C@H]([C@@]([C@@H]8N6C)(C(=O)OC)O)OC(=O)C)CC)OC)C(=O)OC)O</chem>
368	Pallidine	<chem>CN1CC[C@@]23C=C(C(=O)C=C2[C@@H]1CC4=CC(=C(C=C34)OC)O)OC</chem>
369	O-Methylpallidine	<chem>CN1CC[C@@]23C=C(C(=O)C=C2[C@@H]1CC4=CC(=C(C=C34)OC)OC)OC</chem>
370	trigonostemine G	<chem>COC1=CC=C2C(NC3=C2C=CN=C3C(=O)C2=CNC3=C2C=C(O)C=C3)=C1</chem>
371	trigonostemone J	<chem>CC(=O)C(\C1=CNC2=C1C=CC=C2)=C1/C(=O)NC2=CC(O)=CC=C12</chem>
372	Trigonostemine A	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)C4=CNC5=CC=CC=C54</chem>
373	Trigonostemine B	<chem>COC1=CC2=C(C=C1)C3=C(N2)C(=NC=C3)C(=O)C4=CNC5=C4C=CC(=C5)OC</chem>
374	Onosmin B	<chem>[H]N(CC1=CC=C(C)C=C1)C1=CC=CC=C1C(=O)OC</chem>
375	Onosmin A	<chem>[H]OC(=O)C1=CC=CC=C1N([H])CC1=CC=C(C)C=C1</chem>
376	N-methyltryptamine	<chem>[H]C1=C(CNC)C2=CC=CC=C2N1</chem>
377	2-hydroxymethyl-N-methyltryptamine	<chem>CNCC1=C(CO)NC2=CC=CC=C12</chem>

378	N-methyltyramine	<chem>[H]N(C)CCC1=CC=C(O)C=C1</chem>
379	hordenine	<chem>CN(C)CCC1=CC=C(O)C=C1</chem>
380	Julocrotine	<chem>CCC(C)C(=O)NC1CCC(=O)N(C1=O)CCC2=CC=CC=C2</chem>
381	N[N-(2-methylbutanoyl)glutaminoyl]-2-phenylethylamine	<chem>CCC(C)C(=O)NC(CCC(N)=O)C(=O)NCCC1=CC=CC=C1</chem>
382	N-1,N-2,N-3-triisopentenylguanidine	<chem>CC(=CCNC(=NCC=C(C)C)NCC=C(C)C)C</chem>
383	Magnoflorine	<chem>C[N+]1(CCC2=CC(=C(C3=C2[C@@H]1CC4=C3C(=C(C=C4)OC)O)OC)OC)C</chem>
384	Taspine	<chem>CN(C)CCC1=CC(=C2C3=C1C(=O)OC4=C(C=CC(=C34)C(=O)O2)OC)OC</chem>
385	Magnoflorine methine	<chem>COC1=CC=C2CCC3=C(CCN(C)C)C=C(OC)C(O)=C3C2=C1O</chem>
386	taspinic acid	<chem>COC1=CC=C(OO)C(=C1O)C1=C(O)C(OC)=CC(CCN(C)C)=C1OO</chem>
387	Megastachine	<chem>C[C@@H]1C[C@H]2[C@H]3[C@@H]([C@@H]4[C@]25CCCN(C5=C1)CCC4)[C@@H](C(=O)O3)[C@@H](C)O</chem>
388	ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
389	ricicomina A	<chem>CNC(=O)C1=C(OC)C=C(O)C(=C1)C#N</chem>
390	ricinine	<chem>CN1C=CC(=C(C1=O)C#N)OC</chem>
391	N-demethylricinine	<chem>COC1=C(C(=O)NC=C1)C#N</chem>
392	4-[2-formyl-5-(methoxymethyl)-1H-pyrrol-1-yl]butanoic acid	<chem>COCC1=CC=C(N1CCCC(=O)O)C=O</chem>
393	TETRAMETHYLPYRAZINE	<chem>CC1=C(N=C(C(=N1)C)C)C</chem>
394	Jatropham	<chem>CC1=C[C@H](NC1=O)O</chem>

395	5-Hydroxypyrrolidin-2-one	<chem>C1CC(=O)NC1O</chem>
396	Capsaicin	<chem>CC(C)/C=C/CCCCC(=O)NCC1=CC(=C(C=C1)O)OC</chem>
397	Glaucine	<chem>CN1CCC2=CC(=C(C3=C2[C@@H]1CC4=CC(=C(C=C43)OC)OC)OC)OC</chem>
398	Oxoglucine	<chem>COC1=C(C2=C3C(=C1)C=CN=C3C(=O)C4=CC(=C(C=C42)OC)OC)OC</chem>