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

MAYARA DOS SANTOS MAIA

AVALIAÇÃO *IN SILICO* DO POTENCIAL DE ATIVIDADE DE LIGNANAS E NEOLIGNANAS FRENTE A DOENÇAS NEGLIGENCIADAS E NEURODEGENERATIVAS

JOÃO PESSOA – PB
2021

MAYARA DOS SANTOS MAIA

AVALIAÇÃO *IN SILICO* DO POTENCIAL DE ATIVIDADE DE LIGNANAS E NEOLIGNANAS FRENTE A DOENÇAS NEGLIGENCIADAS E NEURODEGENERATIVAS

Tese apresentada à Coordenação do Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos do Centro de Ciências da Saúde da Universidade Federal da Paraíba como requisito para a obtenção do título de Doutora em Produtos Naturais e Sintéticos Bioativos. Área de concentração: Farmacoquímica.

ORIENTADOR: PROF. DR. MARCUS TULLIUS SCOTTI

JOÃO PESSOA – PB

2021

M217a Maia, Mayara dos Santos.
Avaliação in silico do potencial de atividade de lignanas e neolignanas frente a doenças negligenciadas e neurodegenerativas / Mayara dos Santos Maia. - João Pessoa, 2021.
311 f. : il.

Orientação: Marcus Tullius Scotti.
Tese (Doutorado) - UFPB/CCS.

1. Leishmania. 2. Trypanossoma cruzi. 3. Schistossoma mansoni. 4. Alzheimer. 5. Lignanas. 6. Neolignanas. 7. Modelos de predição. 8. Docking molecular. 9. Docking consenso. 10. Simulações de dinâmica molecular. 11. Cálculos de energia livre. I. Scotti, Marcus Tullius. II. Título.

UFPB/BC CDU 616.993.161(043)

MAYARA DOS SANTOS MAIA

AVALIAÇÃO *IN SILICO* DO POTENCIAL DE ATIVIDADE DE LIGNANAS E NEOLIGNANAS FRENTE A DOENÇAS NEGLIGENCIADAS E NEURODEGENERATIVAS

Tese avaliada em 10/ 09/ 21

COMISSÃO EXAMINADORA

Prof. Dr. Marcus Tullius Scotti
PhD em Química Orgânica
Universidade Federal da Paraíba
Orientador

Prof. Dr. Eugene N. Muratov
PhD em Química Orgânica
University North Carolina (UNC/USA)
Professor visitante da UFPB
Avaliador interno

Prof. Dr. Gerd Bruno da Rocha
PhD em Química
Universidade Federal da Paraíba
Avaliador externo

Profa. Dra. Maria de Fátima Vanderlei de Souza
PhD em Farmácia
Universidade Federal da Paraíba
Avaliadora interna

Prof. Dr. Fernando Batista da Costa
Doutor em Química Orgânica
Universidade de São Paulo
Avaliador externo

Dedico este trabalho a Deus, pelo dom da vida e por tão grande amor e misericórdia!

E Maria disse: "Minha alma glorifica ao Senhor, meu espírito exulta de alegria em Deus, meu Salvador, porque olhou para a sua pobre serva.".

(Lucas 1, 46-48)

AGRADECIMENTOS

A Deus, pelo dom da vida, por tão grande amor e misericórdia. A Ele toda honra, toda glória e todo louvor! A Virgem Santíssima, minha Mãe e Rainha, que me conforta e me guia a cada momento, a cada etapa da minha vida.

A minha família, que sempre esteve me apoiando nas decisões por mim tomadas e por todo suporte!

Ao meu orientador, prof. Dr. Marcus Tullius Scotti, primeiro por ter acreditado em mim e segundo por ter contribuído com sua orientação. Deixo meus agradecimentos especiais, pois incentivou as minhas ideias e trabalhos acadêmicos. Reconheço e aprecio o apoio e as orientações dadas durante essa etapa importante da minha vida.

A profa. Dra. Luciana Scotti, pessoa maravilhosa, muito sábia, que também muito me ajudou nessa etapa.

Aos membros da banca avaliadora: Prof. Dra. Fátima Vanderley, Prof. Dr. Eugene Muratov, Prof. Dr. Gerd Rocha e Prof. Dr. Fernando Batista, por aceitarem o convite e contribuir com o meu trabalho.

A Josélia, uma grande amiga e irmã que conheço desde a graduação. Sempre a encontrei nos momentos mais difíceis e desafiadores da minha vida. Foi um grande canal da graça de Deus, pois mesmo sem eu dizer nada, ela me dizia o que eu precisava ouvir.

A Gabriela Cristina, que mais do que uma amiga, é uma irmã que Deus me deu. Foram muitos os desafios nesse tempo, muitos acontecimentos intensos que nos fizeram amadurecer para entendermos que Deus cuida de tudo. Tivemos a ajuda uma da outra para que a graça de Deus nos alcançasse e nos consolasse.

Aos meus amigos de laboratório: Natália Ferreira, Érika Paiva, Renata Barros, Chonny Herrera e Luana.

Aos meus amigos da Comunidade Doce Mãe de Deus, especialmente: Mônica, Karla, Kaline e Kleysse. São muitos os amigos e irmãos da CDMD que não dá pra listar todos aqui, mas deixo meus agradecimentos.

Aos meus amigos da Paróquia Santo Antônio Deus: Ysabelle Rocha, Milena Coutinho, Rodrigo Sena, Felipe Sena, Iris Parente, Renata, Juliana, Jéssica Sobral, Amanda Loyola pela amizade, incentivo, orações e carinho!

A Universidade Federal da Paraíba e ao Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioaticos, onde realizei minha formação acadêmica e a todos que contribuíram nesta etapa importante da minha vida.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

MAIA, Mayara dos Santos. **Avaliação *in silico* do potencial de atividade de lignanas e neolignanas frente a doenças negligenciadas e neurodegenerativas,** 2021.
Tese (Pós-Graduação em Produtos Naturais e Sintéticos Bioativos – Farmacoquímica).
Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, 2021.

RESUMO

Os produtos naturais são considerados potenciais fontes de novos agentes terapêuticos devido à diversidade de estruturas e suas propriedades. Dentre os produtos naturais, destacam-se as lignanas e neolignanas, que possuem uma variedade de atividades biológicas, mas que devido à diversidade estrutural, se faz necessário identificar e investigar novas fontes de efeitos farmacológicos. As doenças negligenciadas ameaçam e atingem milhões de pessoas ao redor do mundo. Dentre elas, a leishmaniose, a doença de Chagas e a esquistossomose, cujo os tratamentos quimioterápicos disponíveis apresentam alta toxicidade e dificilmente apresentam eficácia na fase crônica da doença. As doenças degenerativas também têm acometido diversas pessoas, sendo o Alzheimer mais comum. O tratamento também é limitado, pois não evita a progressão da doença. Variadas abordagens computacionais da Quimio e Bioinformática podem auxiliar, mediar, orientar e identificar novos compostos para o tratamento de diversas doenças. Portanto, o objetivo deste trabalho é avaliar o potencial farmacológico de lignanas e neolignanas frente a doenças negligenciadas e neurodegenerativas com o auxílio de ferramentas e abordagens computacionais. No capítulo 2, foram avaliadas lignanas a partir do banco de dados ChEMBL e aplicado abordagens como perfil farmacocinético, análise combinada baseada no ligante e na estrutura, modelagem por homologia, predição de resistência e simulações de dinâmica molecular. Quatro dentre as lignanas selecionadas na triagem, foram isoladas e testadas contra formas promastigotas de *Leishmania major* e *L. (Viannia) braziliensis*. Os resultados mostraram que o composto mais ativo, o (159) epipinoresinol-4-O-β-D-glucopiranosídeo, apresentou um valor IC_{50} de 5,39 µM para *L. braziliensis* e valor IC_{50} de 36,51 µM para *L. major*. No capítulo 3, previmos o potencial tripanomicida de 47 neolignanas usando modelos preditivos, docking molecular, simulações de dinâmica molecular e cálculos de energia livre. Dos compostos analisados, dois foram isolados e mostraram inibir o crescimento de formas epimastigotas em concentrações de 9,64 e 8,72 µM, e formas tripomastigotas em 4,88 e 2,73 µM. Enquanto que no capítulo 4, abordagens *in silico*, usando análise de perfil farmacocinético, docking consenso, modelos preditivos consenso, simulações de dinâmica molecular e cálculos de energia livre também foram utilizados para selecionar lignanas potenciais e seletivas contra um importante alvo do *Schistossoma mansoni*. Quatro lignanas obtiveram excelentes resultados e sugerimos ser um alternativa terapêutica em casos de resistência. No capítulo 5, lignanas foram analisadas com o objetivo de identificar compostos potenciais e *multi-target* para o tratamento do Alzheimer. Uma análise combinada, com base no ligante e na estrutura, seguida pela previsão das propriedades de absorção, distribuição, metabolismo, excreção e toxicidade (ADMET) foi realizada. Os resultados mostraram que a análise combinada foi capaz de selecionar 139 lignanas potencialmente ativas e multitarget, conferirindo alternativas de tratamento através da atividade neuroprotetiva e antioxidante. O capítulo 6 é uma revisão que descreve vários estudos, abordagens e métodos de docking consenso.

Palavras-chave: *Leishmania*, *Trypanossoma cruzi*, *Schistossoma mansoni*, Alzheimer, lignanas, neolignanas, modelos de predição, docking molecular, docking consenso, simulações de dinâmica molecular, cálculos de energia livre.

MAIA, Mayara dos Santos**. In silico evaluation of the activity potential of lignans and neolignans against neglected and neurodegenerative diseases**, 2021.
Thesis (Post-Graduation in Natural and Synthetic Bioactive Products – Pharmacochemistry).
Health Sciences Center, Federal University of Paraíba, João Pessoa, 2021.

ABSTRACT

Natural products are considered potential sources of new therapeutic agents due to the diversity of structures and their properties. Among the natural products, lignans and neolignans stand out, which have a variety of biological activities, but due to their structural diversity, it is necessary to identify and investigate new sources of pharmacological effects. Neglected diseases threaten and affect millions of people around the world. Among them, leishmaniasis, Chagas disease and schistosomiasis, whose available chemotherapy treatments are highly toxic and hardly effective in the chronic phase of the disease. Degenerative diseases have also affected many people, with Alzheimer's being the most common. Treatment is also limited as it does not prevent disease progression. Several computational approaches in Chemo and Bioinformatics can help, mediate, guide and identify new compounds for the treatment of various diseases. Therefore, the objective of this work is to evaluate the pharmacological potential of lignans and neolignans against neglected and neurodegenerative diseases with the help of several computational tools and approaches. In chapter 2, lignans were evaluated from the ChEMBL database and applied approaches such as pharmacokinetic profiling, combined analysis based on ligand and structure, homology modeling, resistance prediction and molecular dynamics simulations. Four of the lignans selected in the screening were isolated and tested against promastigote forms of *Leishmania major* and *L. (Viannia) braziliensis*. The results showed that the most active compound, (159) epipinoresinol-4-O-β-D-glucopyranoside, had an IC_{50} value of 5.39 µM for *L. braziliensis* and an IC_{50} value of 36.51 µM for *L. major*. In Chapter 3, we predicted the trypanocidal potential of 47 neolignans using predictive models, molecular docking, molecular dynamics simulations, and free energy calculations. Of the compounds analyzed, two were isolated and showed to inhibit the growth of epimastigote forms at concentrations of 9.64 and 8.72 µM, and trypomastigote forms at 4.88 and 2.73 µM. While in Chapter 4, in silico approaches using pharmacokinetic profile analysis, consensus docking, consensus predictive models, molecular dynamics simulations and free energy calculations were also used to select potential and selective lignans against an important target of *Schistosoma mansoni*. Four lignans had excellent results and we suggest that they are a therapeutic alternative in cases of resistance. In chapter 5, lignans were analyzed with the aim of identifying potential and multitarget compounds for the treatment of Alzheimer's. A combined analysis, based on ligand and structure, followed by prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties was performed. The results showed that the combined analysis was able to select 139 potentially active and multitarget lignans, providing treatment alternatives through neuroprotective and antioxidant activity. Chapter 6 is a review that describes various studies, approaches, and methods of consensus docking.

Keywords: *Leishmania*, *Trypanosoma cruzi*, *Schistossoma mansoni*, Alzheimer, lignans, neolignans, prediction models, molecular docking, consensus docking, molecular dynamics simulations, free energy calculations.

LISTAS DE ABREVIARURAS, SIGLAS E FÓRMULAS

AA	Arachidonic acid (Ácido aracdônico)
AD	Alzheimer's disease (Doença de Alzheimer)
AD4	AutoDock 4
ADMET	Absorção, distribuição, metabolismo, excreção e toxicidade
Ala	Alanina
AMPK2	α2 Subunit of monophosphate-activated protein kinase (Subunidade α2 da proteína quinase ativada por monofosfato)
APP	β-Amyloid precursor protein (Proteína precursora de β-amilóide)
Arg	Arginina
Asn	Asparagina
Asp	Ácido aspártico
AUC	Area under the curve (Área sob a curva)
BACE1	Beta-secretase 1
BNL	Brookhaven National Laboratories
BPPSA	N-(2-bromophenyl)pyrrolidine-1-sulfonamide (N- (2-bromofenil) pirrolidina-1-sulfonamida)
CADD	Computer Aided Drug Design (Projeto de drogas auxiliado por computador)
CAIII	Carbonic Anhydrase III (Anidrase carbônica III)
CASP	Critical Assessment of Protein Structure Prediction (Avaliação crítica da previsão da estrutura da proteína)
CD	Chagas disease (Doença de Chagas)
CE_{50}	Concentração da metade máxima efetiva
ChEBI	Chemical Entities of Biological Interest (Laboratório Europeu de Biologi- Molecular)
CL	Cutaneous leishmaniasis (Leishmaniose cutânea)
CNS	Central Nervous System (Sistema Nervoso Central)
CO_2	Dióxido de carbono
COX-2	Cycloxygenase 2
CR	Rank of averages (Classificação das médias)
CYP	Citocromo
Cys	Cisteína

DCPIP	Dichlorophenolindophenol (Diclorofenolindofenol)
DHODH	Dihydroorotate dehydrogenase (Diidroorotato desidrogenase)
DIR	Proteína dirigente
DM	Dinâmica molecular
DMSO	Dimetilsulfóxido
DN	Doenças negligenciadas
DNA	ácido desoxirribonucleico
DPP-4	Dipeptidyl peptidase-4 (Dipeptidil peptidase-4)
DYNLL1	Dynein light chain 1 (Cadeia leve de dineína 1)
EGFR	Epidermal growth factor (Fator de crescimento epidérmico)
EtOAc	Ethyl acetate (Acetato de etila)
FAAH	Fatty acid amide hydrolase (Amida hidrolase de ácido graxo)
FBDD	Fragment-Based Lead Discovery (Descoberta de lead baseada em fragmento)
FBS	Fetal bovine serum (Soro fetal bovino)
FDA	Food and Drug Administration (Administração de Alimentos e Medicamentos)
FEP	Free energy perturbation (Perturbação de energia grátis)
FLAP	Fingerprints for Ligands and Proteins (Impressões digitais para ligantes e proteínas)
FN	Falso negativo
FP	Falso positivo
FRET	Fluorescence Resonance Energy Transfer (Transferência de energia de ressonância de fluorescência)
fXa	Coagulation enzyme factor Xa (Fator Xa da enzima de coagulação)
FXR	Farnesoid X receptor (Receptor farnesóide X)
GC2	Grand Challenge 2016
GCL	Glutamate-cysteine ligase (Glutamato-cisteína ligase)
GI	Gastrointestinal absorption (Absorção gastrointestinal)
Gln	Ácido glutâmico
Glu	Glutamina
Gly	Glicina
GPDH	Glycerol-3-phosphate dehydrogenase (Glicerol-3-fosfato desidrogenase)
GROMOS	Groningen molecular simulation (Simulação Molecular de Groningen)

GSH	Glutathione synthesis (Síntese de glutationa)
GWI	Gulf War Illness (Doença da Guerra do Golfo)
H1N1	Influenza A subtype (Subtipo Influenza A)
HA	Hemaglutinin (Hemaglutinina)
His	Histidina
HiT QSAR	Hierarchical QSAR tecnology
HRMS	High-resolution mass spectrometry (Espectrometria de massa de alta resolução)
*h*SULT	human Sulfotransferase (Sulfotransferase humana)
IC_{50}	Metade máxima de concentração inibitória
IL	Interleukin (Interleucina)
IL-1Ra	Interleukin 1 receptor antagonist (Antagonista do receptor de interleucina 1)
Ile	Isoleucina
*i*NOS	inducible nitric oxide synthase (Óxido nítrico sintase induzível)
IRHOM2	Rhomboid protease 2 (Protease romboide 2)
IUPAC	International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada)
JNK-3	c-Jun N-terminal kinase 3
Kd	Constante de dissociação
Keap1	Kelch-like ECH-associated protein 1 (Proteína 1 associada a ECH semelhante a Kelch)
Ki	Constante inibitória
LBVS	Ligand-based virtual screening (Triagem virtual baseada no ligante)
Leu	Leucina
LIE	Linear interaction energies (Energias de interação linear)
logBB	Blood–brain barrier (Barreira hematoencefálica)
LogP	Partition coefficient (P) (Coeficiente de partição P)
LRGT	Laciriresinol glicosiltransferase
MADM	Mean absolute deviation from median (Desvio médio absoluto da mediana)
MAP	Mitogen-activated protein (Proteína ativada por mitogênio)
MAPK2	Mitogen-activated protein kinase (Proteína kinase ativada por mitogênio
MbtI	Salicylate synthase of *Mycobacterium tuberculosis*
MCC	Mathews correlation coefficient (Coeficiente de correlação de Mathews)
MCL	Mucocutaneous leishmaniasis (Leishmaniose mucocutânea)

MD	Molecular dynamics (Dinâmica molecular)
Met	Metionina
MIFs	Molecular interaction of fields (Interação molecular de campos)
ML	Machine learning (Aprendizado de máquina)
MM-PBSA	Molecular mechanics Poisson−Boltzmann surface area (Mecânica molecular área de superfície de Poisson-Boltzmann)
MOMT	Matairesinol O-metiltransferase
MT	Metacyclic trypomastigote (Tripomastigota metacíclico)
Mtb-DHFR	Dihydrofolate reductase of *Mycobacterium tuberculosis*
mTOR	Mammalian Target of Rapamycin (Alvo Rapamicina de mamífero)
MVD	Molegro Virtual Docker
N	Negativo
NA	Neuraminidase
NADPH	Nicotinamida adenina dinucleótido fosfato
NCI	National Cancer Institute (Instituto Nacional do Câncer)
NO	Nitric oxide (Óxido nítrico)
NOX1	Nicotinamide Adenine Dinucleotide Phosphate Oxidase 1
NP	Nucleoprotein (Nucleoproteína)
NPT	Constant pressure numerical particles and temperature (Partículas numéricas de pressão constante e temperatura)
NPV	Negative Predictive Value
NPV	Negative predicted value (Valor predito negativo)
NQO1	NADPH Quinone Oxidoreductase 1
Nrf2	Nuclear factor erythroid 2-related (Fator nuclear 2 relacionado ao eritroide)
NVT	Constant number of particles, volume and temperature (Número constante de partículas, volume e temperatura)
OMS	Organização mundial da saúde
OPLS	Optimized potential for liquid simulations (Potencial otimizado para simulações líquidas)
OXA	Oxamniquine
P	Positivo
PA	Acidic polymerase protease (Protease de polimerase ácida)
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
PCA	Principal Component Analysis (Análise de componentes principais)

PDB	Protein Data Bank (Banco de dados de proteínas)
PDE5	Phosphodiesterase 5
PfFabD	Malonyl CoA-acyl carrier protein transacylase of Plasmodium falciparum
P-gp	P-glicoproteína
PGs	Prostaglandins (Prostaglandinas)
Phe	Fenilanina
pIC_{50}	logarithm of 50% inhibitory concentration (logaritmo de concentração inibitória de 50%)
PKDL	Post-kala azal dermal leishmaniasis (Leishmaniose dérmica pós-kalazar)
PLR	Pinoresinol/lariciresinol redutase
PLS	Partial Least Squares (Mínimos Quadrados Parciais)
PN	Produtos naturais
PNGT	Putativa pinoresinol glicosiltransferase
PPV	Positive predicted value (Valor predito positivo)
Pro	Prolina
PSS	(+)-sesamiina por piperitol/sesamina
PTP1B	Protein Tyrosine Phosphatase 1B
PTR1	Pteridine reductase 1 (Pteridina reductase 1)
PUFAs	Polyunsaturated fatty acids (Ácidos graxos poliinsaturados)
PZQ	Praquizantel
QSAR	Quantitative structure–activity relationship (Relação quantitativa estrutura-atividade)
RF	Random Forest (Floresta aleatória)
RMN	Ressonância Magnética Nuclear
RMSD	Root-mean-square deviation (Desvio da raiz quadrada média)
RMSF	Root-mean-square fluctuation (Flutuação quadrática média)
ROC	Receiver operating characteristic (Características operacionais do receptor)
ROS	Reactive oxigen species (Espécies reativas de oxigênio)
SAR	Sructure activity relationships (Relação estrutura atividade)
SASA	Solvent accessible surface area (Área de superfície acessível a solvente)
SBP	Structure-based pharmacophore (Farmacóforo baseado em estrutura)
SBVS	Structure-based virtual screening (Triagem virtual baseada na estrutura)
SC	Scoring consensus (Pontuação consenso)

SDC	Standard Deviation Consensus (Consenso de Desvio Padrão)
Ser	Serina
SGLT	Sodium-glucose cotransporter (Cotransportador sódio-glicose)
SIR	Secoisolariciresinol desidrogenase
SIRGT	Secoisolariciresinol glicosiltransferase
SISGEN	National Management System of Genetic Patrimony and Associated Traditinal Knowledge (Sistema Nacional de Gestão do Patrimônio Genético e Conhecimentos Tradicionais Associados)
SMGT	Sesaminol 2-O-glucosiltransferase
SNP	Single nucleotide polymorphisms (Polimorfismos de nucleotídeo único)
SOD	Superoxide dismutase (Superóxido dismutação)
SULT	Sulfotransferase
TFP	Taxa falso positivo
TgFabD	Malonyl CoA-acyl carrier protein transacylase of Toxoplasma gondii
Thr	Treonina
TJOMT	O-metiltransferase de thujaplicatin
TNF-α	Tumor necrosis factor α (Fator de necrose tumoral α)
TPSA	Topologic Polar Surface Area (Área de superfície polar topológica)
TPV	Taxa verdaeiro positivo
TR	Trypanothione reductase (Tripanothione reductase)
Trp	Triptofano
Tyr	Tirosina
UGPase	UDP-glucose pyrophosphorylase (UDP-glucose pirofosforilase
Val	Valina
VL	Visceral leishmaniasis (Leishmaniose visceral)
VN	Verdadeiro negativo
VP	Verdadeiro positivo
VS	Virtual screening (Triagem virtual)
vSDC	variable Standard Deviation Consensus (Consenso de desvio padrão variável)

LISTA DE FIGURAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3

CAPÍTULO 4

CAPÍTULO 5

CAPÍTULO 6

LISTA DE TABELAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3

CAPÍTULO 4

CAPÍTULO 5

CAPÍTULO 6

SUMÁRIO

Capítulo 1

1. INTRODUÇÃO

O Brasil é o país com maior biodiversidade do mundo, possuindo mais de 50.000 espécies de plantas superiores já descritas e cerca de 25% dos medicamentos são originários de produtos naturais (PNs) obtidos direta ou indiretamente de plantas [1]. Os PNs oferecem características especiais em comparação com moléculas sintéticas convencionais, conferindo vantagens para o processo de descoberta de drogas [2]. Eles são continuamente explorados no desenvolvimento de novos bioativos com diversas aplicações devido às suas estruturas semelhantes a farmacóforo, com propriedades farmacocinéticas e espaço químico único [3]. Devido a essas propriedades, os PNs são uma importante fonte de drogas orais [4].

Otto Richard Gottlieb foi pioneiro no estudo da fitoquímica no Brasil e foi responsável por inúmeros trabalhos na química de produtos naturais, pesquisa de evolução botânica e ecologia química [5]. Ele também contribuiu para o desenvolvimento de drogas com neolignanas, a partir de um vasto estudo sobre as neolignanas e lignanas [6].

As lignanas e neolignanas são uma classe de produtos naturais com um amplo espectro de importantes atividades biológicas [7]. Lignanas e neolignanas são metabólitos secundários derivados do acoplamento oxidativo de fenilpropanóides e as atividades biológicas relatadas variam entre antioxidante, antitumoral, antiinflamatório, antineurodegenerativo antimicrobiano, entre outros [8]. Segundo Gordaliza *et al* (2004) [9], as lignanas e neolignanas podem ser encontradas em mais de 60 famílias de plantas vasculares e foram isoladas de diferentes partes da planta. Devido à existência de diversas subclasses de lignanas e neolignanas, faz-se necessário a exploração de novas estruturas e efeitos farmacológicos.

As Doenças Negligenciadas (DNs), segundo a Organização Mundial da Saúde (OMS), corresponde a um grupo de doenças que atingem principalmente populações pobres [10]. A OMS tem recomendado uma série de estratégias que visam controlar, prevenir e possivelmente eliminar várias DN. Porém, as DNs estão prejudicando a vida de mais de um bilhão de pessoas ao redor do mundo e ameaçam a vida de milhões de pessoas. As doenças classificadas como negligenciadas pela OMS são causadas por diversos agentes etiológicos, três delas são comuns no Brasil e representam um problema de saúde pública: a leishmaniose, a doença de Chagas e a esquistossomose [11].

A leishmaniose é causada por protozoários do gênero *Leishmania* e ocorre com prevalência em 88 países [12], [13]. O Brasil está entre os três países que possuem 75% dos casos da doença no mundo. As manifestações clínicas podem evoluir para a fase crônica, podendo levar a morte. Além disso, a leishmaniose visceral é a segunda maior causa de mortalidade no Brasil dentre as doenças parasitárias [14].

A doença de Chagas é causada pelo protozoário *Trypanosoma cruzi* que leva às maiores taxas de morbidade e mortalidade na América Latina do que qualquer outro parasita, resultando em diminuições significativas na qualidade de vida [15]. O Brasil é um dos três países com maior número de casos estimados de infecção. As manifestações clínicas crônicas ocorrem na minoria, mas a fase aguda pode perdurar por muito tempo, prejudicando o indivíduo [16], [17].

A esquistossomose é causada por platelmintos da classe Trematoda [18] e representa uma das principais causas de morbidade no mundo, perdendo apenas para a malária como uma das principais doenças infecciosas [19]. Estima-se que a doença afete mais de 250 milhões de pessoas em 78 países do mundo e seja responsável por cerca de 280.000 mortes a cada ano [19]. Os sintomas podem levar a complicações na fase crônica podendo levar a óbito [20].

O tratamento para a leishmaniose, a doença de Chagas e a esquistossomose é à base de quimioterápicos que têm mostrado em muitos casos ser ineficiente ou prejudicial, devido a resistência a droga e aos efeitos colaterais por causa do alto índice de toxicidade, como tem sido abordado em diversos estudos [21], [22], [23]. Além disso, não existem vacinas disponíveis, embora haja algumas vacinas candidatas para o tratamento dessas doenças que se encontram na fase de testes pré-clínicos e clínicos [24].

As doenças neurodegenerativas (DNs) são distúrbios caracterizados pela perda progressiva de neurônios seletivamente vulneráveis, o que leva à perda neuronal estática, associada à deposição de proteínas com propriedades físico-químicas alteradas no cérebro e em órgãos periféricos [25], [26]. As DNs prejudicam a memória, a cognição e o movimento; mas atualmente nenhuma delas é curável e os tratamentos disponíveis apenas controlam os sintomas ou interrompem a progressão da doença [27]. De todas as DNs, o Alzheimer tem sido a mais comum e mais estudada [28].

A disponibilidade do sequenciamento e análises proteômicas forneceu macromoléculas e vias que podem ser alvos potenciais contra as DNs e doenças neurodegenerativas [29]. Além disso, diversas ferramentas computacionais contribuem para a criação de bases de dados, previsão da função das proteínas, modelagem de estruturas proteicas, simulação da cinética de vias metabólicas, previsão da toxicidade, observação da interação entre alvos e ligantes, predizer a atividade biológica, entre outros; o que proporciona o planejamento e identificação de fármacos com potencial para o tratamento de diversas doenças. Dessa forma, aumenta a chance de potencialidade, eficácia e baixa toxicidade de possíveis fármacos [30], [31].

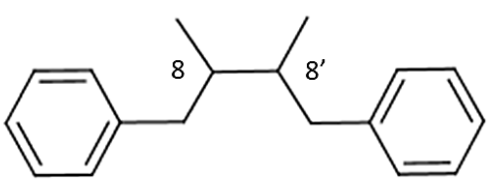
Nesse contexto, a identificação de produtos naturais com diversidade estrutural representa possibilidades atraentes para o desenvolvimento de fármacos com variadas atividades biológicas. Portanto, o presente trabalho buscou avaliar o potencial farmacológico de lignanas e neolignanas frente a doenças negligenciadas e neurodegenerativas com o auxílio de ferramentas computacionais da Quimio e Bioinformática, bem como a validação experimental.

2. FUNDAMENTAÇÃO TEÓRICA

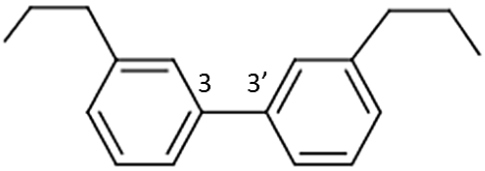
2.1 Lignanas, neolignanas e seu potencial farmacológico

As lignanas e neolignanas são tradicionalmente definidas como uma classe de metabólitos secundários derivados da dimerização oxidativa de duas ou mais unidades fenilpropanóides [32]. A forma como estas unidades são ligadas entre si determinam sua classificação: as lignanas ligam-se pela posição 8 e 8' da cadeia alifática, enquanto que as neolignanas são geradas pela ligação de qualquer outra posição da unidade fenilpropanoide que não contém a ligação 8 e 8' [33] (Figura 1).

Figura 1. Esqueletos carbônicos de lignanas e neolignanas.



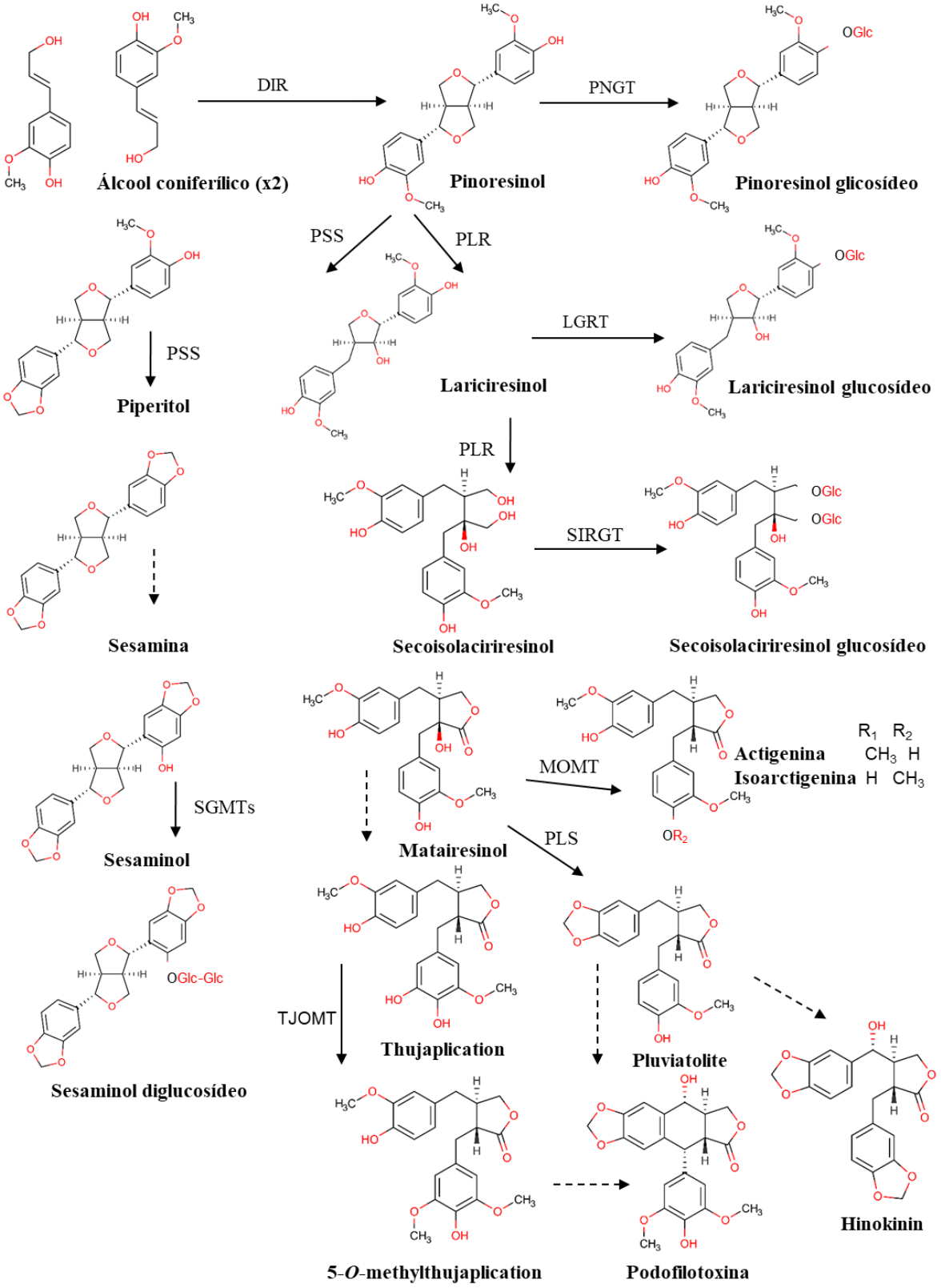
Lignana



3,3'-Neolignana

Lignanas e neolignanas são metabólitos originários da via biossintética do ácido chiquímico, que se origina do acoplamento do álcool aquiral *E*-coniferílico, levando à geração de pinoresinol, uma lignana basal [8] (Figura 2). Embora uma pinoresinol sintase ainda não tenha sido identificada, uma proteína dirigente demonstrou participar da dimerização estereoespecífica do álcool *E*-coniferílico [34]. A biossíntese de álcoois coniferílicos é iniciada com desaminação da fenilalanina pela fenilalanina amonialilase para formar o ácido cinâmico, que é hidroxilado por uma enzima P450, para formar o ácido *p*-cumarico [35]. O álcool coniferílico é ainda mais reduzido na presença de uma molécula de nicotinamida adenina dinucleótido fosfato (NADPH). A formação de um éster facilita o primeiro passo de redução, introduzindo melhor um grupo de saída para a reação dependente [35]. Em diversas espécies vegetais, o pinoresinol é reduzido em lariciresinol e, em seguida, secoisolariciresinol [34]. O pinoresinol também sofre glicosilação através da glicosiltranferase dependente de UDP-glicose.

Figura 2. Vias biossintéticas das principais lignanas catalisadas por enzimas identificadas e não identificadas, respectivamente.



DIR – Proteína dirigente
PSS – (+)-sesamiina por piperitol/sesamina
PLR – pinoresinol/lariciresinol redutase
SIR – secoisolariciresinol desidrogenase
PNGT – putativa pinoresinol glicosiltransferase
LRGT – laciriresinol glicosiltransferase
SIRGT – secoisolariciresinol glicosiltransferase
MOMT – matairesinol O-metiltransferase
TJOMT – O-metiltransferase de thujaplicatin
SMGT – sesaminol 2-O-glucosiltransferase

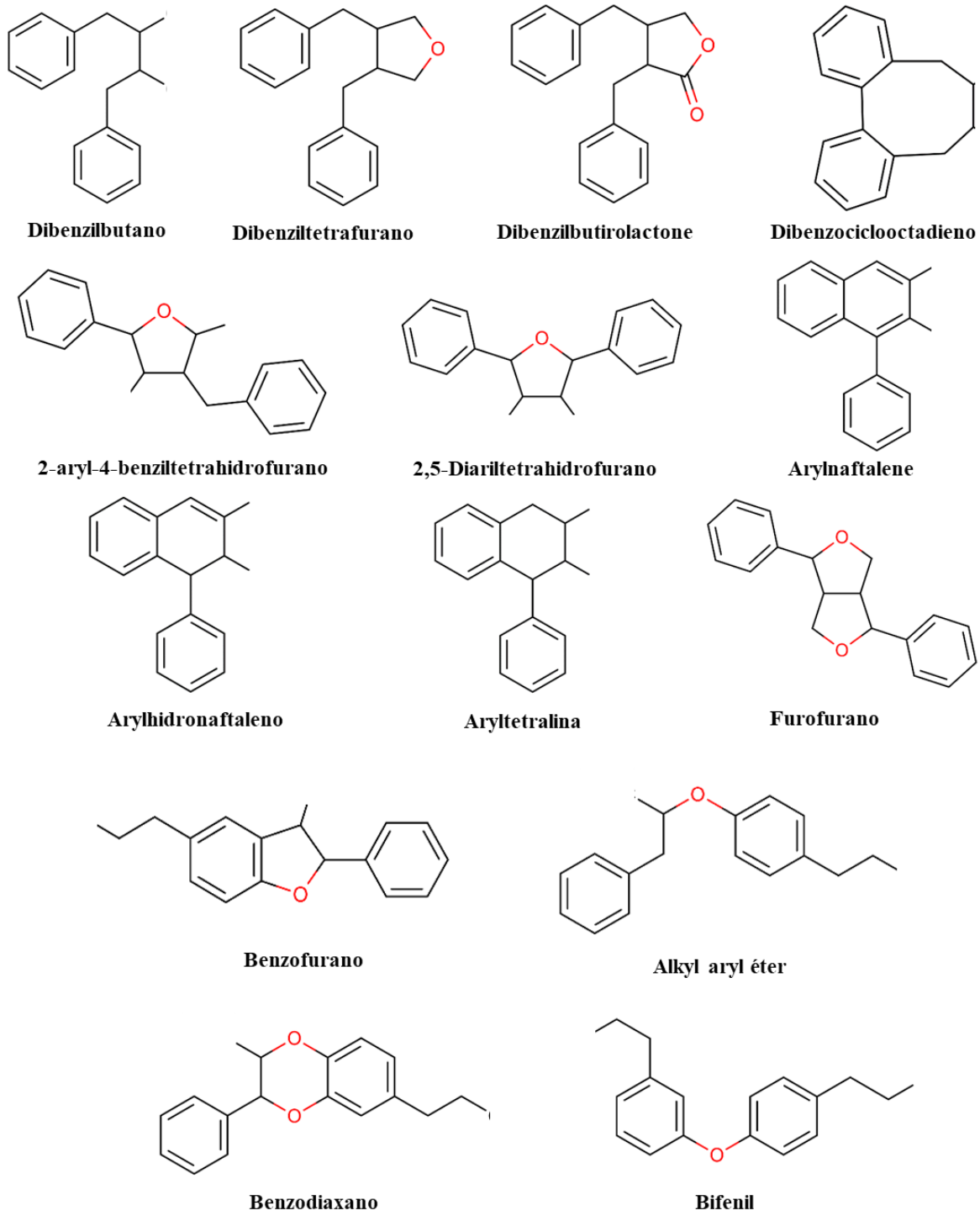
Fonte: Teponno; Kusari; Spiteller [5].

A conversão de álcool coniferílico em matairesinol foi demonstrado em várias espécies de plantas. O matairesinol é metabolizado em arctigenina/isoarctigenina via metilação de um grupo hidroxila fenólico. Além disso, nas plantas de *Linum*, *Anthriscus* e *Podophyllum*, o matairesinol também é convertido em hinokinin, yatein ou podophyllotoxin através de múltiplas vias biossintéticas [35]. Enquanto que a tujaplicatina é precursor biossintético da podofilotoxina. O pinoresinol também pode ser metabolizado em piperitol, seguidos de posterior conversão em (+) – sesamina. O sesaminol é glicosilado na sua posição 2-hidroxila pelo sesaminol 2-O-glucosiltransferase (SMGT) em *Sesamum radiatum* (Schum. and Thonn.), *Sesamum indicium* (Sesame) e *Sesamum alatum* Thonn [35]. Lariciresinol e secoisolariciresinol também podem ser glicosilados pela ação de enzimas específicas, assim como o pinoresinol.

As lignanas são classificadas em dez subclasses com base em seus padrões estruturais, incluindo: seus esqueletos de carbono, a maneira pela qual o oxigênio é incorporado nos esqueletos e o padrão de ciclização. As subclasses são: dibenzilbutano, 3,4-dibenziltetrahidrofurano, 2,5-diariltetrahidrofurano, 2-aril-4-benziltetrahidrofurano, dibenzilbutirolactone, dibenzilciclooctadieno, arilnaftaleno, arilhidronaftaleno, ariltetralina e furofurano [8] (Figura 3). As neolignanas, por sua vez, têm estruturas mais variadas, contendo 15 subclasses sendo os mais comuns benzofuranos, 1,4-benzodioxanos, éteres alquilarílicos, bifenilos, ciclobutanos, 8-l-biciclo octanos, 8-3 biciclo octanos e éteres bifenílicos [36] (Figura 3).

Em geral, os dímeros de fenilpropanóides possuem uma diversidade de propriedades biológicas interessantes, tornando-os uma fonte importante de novos candidatos a medicamentos e/ou bases estruturais exploráveis no campo da química medicinal [8]. Foram encontrados em mais de 70 famílias de plantas e exibem várias atividades biológicas significativas e potentes; incluindo atividade antioxidante [37], [38], anti-inflamatória [39], hepatoprotetora [40], antitumoral [38], [41], antimicrobiana [8], tripanomicida [42], [43], neuroprotetora [44] e larvicida [45].

Figura 3. Subclasses estruturais de lignanas e principais classes de neolignanas.



Um estudo realizado por Pilkington (2018) [36] com o objetivo de avaliar perfis ADMET (absorção, distribuição, metabolismo, excreção e toxicidade) das lignanas e estabelecer semelhanças de compostos à fármacos (*druglikness*), mostrou que mais de 75% das lignanas são consideradas *druglikness*. A pesquisa também realizou uma análise quimiométrica para investigar a relação entre os descritores moleculares estudados e as

subclasses de lignanas e observaram que as flavolignanas e as lignanas conjugadas a carboidratos são menos semelhantes a drogas. O estudo conclui que as lignanas mostram um nível alto de semelhança medicamentosa. A alta proporção de semelhança medicamentosa de lignanas clássicas e neolignanas permite a justificabilidade e a importância de investigar lignanas como possíveis drogas.

As plantas têm sido usadas há séculos na medicina tradicional e muitos dos medicamentos prescritos em países industrializados contêm compostos derivados de plantas, direta ou indiretamente via semissíntese [46]. A podofilotoxina (PTOX) é uma lignana que foi isolada como medicamento antitumoral ativo das raízes de espécies do gênero *Podophyllum* e mais recentemente do gênero *Linum*. Etoposídeo, teniposídeo e etofos são derivados semi-sintéticos de podofilotoxina e também são usados no tratamento do câncer [47]. Recentemente, Nikolov *et al*. (2007) [48], demonstraram a atividade citotóxica de extratos de *Linum tauricum* Willd, *L. leonii* F.W.Schultz, e *L. narbonense* (Blue Flax), contendo altas concentrações de lignanas semelhantes a PTOX, que podem ser uma nova fonte de medicamentos à base de lignanas. Esses estudos revelam a importância de investigar as lignanas e neolignanas como futuros medicamentos.

2.2 A Quimioinformática e a triagem virtual

A quimioinformática é uma ciência interdisciplinar que utiliza recursos das ciências da computação e informação para resolver problemas da química [49]. Consiste em aspectos práticos e técnicas de representação, manipulação e processamento de estruturas químicas; exploração de bases de dados químicos e descoberta de novos compostos com atividade e/ou propriedades desejadas [50].

A busca virtual é uma aplicação da quimioinformática que consiste em pré-selecionar os compostos com o auxílio do computador a partir de bancos de dados virtuais com um grande número de moléculas [31], [51]. Desta forma, é possível otimizar e selecionar os compostos que deverão ser submetidos aos testes de atividade *in vitro* e *in vivo*. Esta pré-seleção é feita triando os compostos de interesse por meio da triagem virtual baseada na estrutura (do inglês *Structure-based virtual screening - SBVS*) ou triagem virtual baseada no ligante (do inglês *Ligand-based virtual screening - LBVS*) [52], [51].

O projeto de drogas auxiliado por computador (do inglês "*Computer Aided Drug Design*" – CADD) consiste em ferramentas poderosas que podem reduzir o número de ligantes que precisam ser selecionados em ensaios experimentais e que têm contribuído na descoberta de muitas drogas farmacêuticas disponíveis que obtiveram a aprovação da FDA [53]. São exemplos, os medicamentos saquinavir, raltegravir e amprenavir, usados para o tratamento de infecções por HIV; o dorzolamida, usado para o tratamento de glaucoma; zolmitriptano: usado no tratamento da enxaqueca; norfloxacina, usado em infecções do trato urinário e o losartan, usado para tratar a hipertensão [53].

Nos subtópicos sequentes são apresentados algumas bases de dados, procedimentos, técnicas e metodologias que abrangem a quimioinformática e a triagem virtual.

2.2.1 Base de dados de compostos químicos

A descoberta racional de fármacos através de métodos computacionais inicia-se com a escolha do banco de dados de compostos, ou série de compostos e de alvos. Após essa etapa, é possível aplicar a triagem virtual. Atualmente, existem várias bases de dados de compostos químicos que trazem diversas informações relevantes para as mais diversas áreas da química e que podem ser usados para a construção de modelos e predição de atividade biológica [50].

Com o avanço de pesquisas através do uso de diversos bancos de compostos, houve uma melhoria na eficiência da triagem virtual para descoberta de drogas. Isso porque bancos de dados químicos diversificados não enriquecem apenas uma seleção de compostos que possuem um alta probabilidade de ser ativo, mas também apresentam mais opções para atender à demanda dos usuários, como rastreio similar a medicamentos, mapeamento de farmacóforos e triagem baseada em ligantes [54].

Existem seis tipos de bases químicas públicas: 1) Base de dados de informações químicas - esse tipo de banco de dados é projetado especificamente para armazenar informações químicas, incluindo estruturas de cristal, espectros, reações e sínteses e dados termofísicos; 2) banco de dados de bioatividade - fornece informações sobre dados de bioatividade, como constante inibitória (Ki), constante de dissociação (Kd), metade máxima na concentração inibitória (IC_{50}), metade máxima efetiva concentração (CE_{50}), entre outros; 3) banco de dados de medicamentos - é um recurso exclusivo de bioinformática e quiminformática que combina dados detalhados de medicamentos com informações

abrangentes sobre o alvo das drogas; 4) banco de produtos naturais - contém as informações sobre vias, estruturas e outros associados à síntese e degradação de produtos naturais; 5) banco de dados disponível - contém compostos disponíveis oferecidos por fornecedores de produtos químicos e 6) banco de dados de fragmentos - consiste em dados de estrutura e propriedade de fragmentos como informações físico-químicas e preferências vinculativas do site [54]. A tabela 1 mostra informações sobre algumas bases de dados. Bancos de dados de patentes também foram incluídos.

Tabela 1. Bases de dados públicos de estruturas químicas.

Base de dados	Instituto	Quantidade de compostos	Site	Ref.
Informação química				
ChemSpider	Sociedade Royal de Química	59.000.000	http://www.chemspider.com	[55]
ChEBI	Laboratório Europeu de Biologia Molecular	51.809	http://www.ebi.ac.uk/chebi/	[56]
Bioatividade				
PubChem	Centro Nacional de Informação Biotecnológica	91.577.678	https://pubchem.ncbi.nlm.nih.gov/	[57]
ChEMBL	Laboratório Europeu de Biologia Molecular	2.101.843	https://www.ebi.ac.uk/chembl/	[58]
BindingDB	Universidade da California	1.346.745	http://www.bindingdb.org	[59]
ChemBank	Instituto Amplo de Harvard e MIT	800.000	http://chembank.broadinstitute.org	[60]
PDBbind	Instituto de Química Orgânica de Xangai	16.179	http://www.pdbbind.org.cn	[61]
Medicamentos				
DrugBank	Universidade de Alberta e Centro de Inovação Metabolômica	8.261	https://www.drugbank.ca	[62]
Produtos naturais				
Super Natural II	Universidade de Charite	325.509	http://bioinformatics.charite.de/supernatural	[63]
SistematX	Universidade Federal da Paraíba	2.150	http://sistematx.ufpb.br	[64]
Universal Natural	Universidade de	229.358	http://pkuxxj.pku.edu.cn/UNPD/	[65]

Product	Peking		index.php	
CamMedNP	Universidade de Martin-Luther	126.140		[66]
TCM Database@ Taiwan	Laboratório de Biologia Computacional e Sistema	37.170	http://tcm.cmu.edu.tw	[67]
Disponibilidade química				
ZINC	Universidade da Califórnia em São Francisco	34.000.000	http://zinc.docking.org/	[68]
NCI	Laboratório Nacional Frederick para pesquisa sobre câncer	500.000	https://cactus.nci.nih.gov/	[69]
Fragmentos				
FDB-17	Universidade do Bem	10.101.204	http://gdb.unibe.ch/downloads/	[70]
Fragment-Store	Universidade de Medicina de Berlim	162.256	http://bioinfapplied charite.de/fragment_store/	[71]
PADFrag	Universidade Central China Normal	5.919	http://chemyang.ccnu.edu.cn/ccb/ database/PA DFrag/	[72]
Banco de dados de patentes				
USPTO	Escritório de patentes e marcas registradas dos estados unidos	-	https://www.uspto.gov/patents/search	
Espacenet		130.000.000	https://www.epo.org/searching-for-patents/technical/espacenet.html	

Adpaptado de Yang *et al*, (2018) [54].

2.2.2 Relação quantitativa estrutura-atividade

A modelagem da relação quantitativa estrutura-atividade é uma das principais ferramentas computacionais empregadas na química medicinal [73]. Um modelo quantitativo QSAR (do inglês "*Quantitative structure–activity relationship*" – QSAR) é representado por meio de uma equação matemática que relaciona as propriedades dos compostos investigados com suas atividades biológicas e que possui significância estatística [74]. Os modelos de relação estrutura-atividade quantitativa são particularmente úteis para diversos estudos sobre toxicidade, mutagenicidade, carcinogenicidade, desenho de medicamentos e predição de atividade biológica [75]. No entanto, nos últimos anos, a modelagem QSAR encontrou aplicações mais amplas na descoberta medicamentos por meio de triagem virtual, bem como na área de previsão de propriedades semelhantes a drogas e avaliação de risco químico [76].

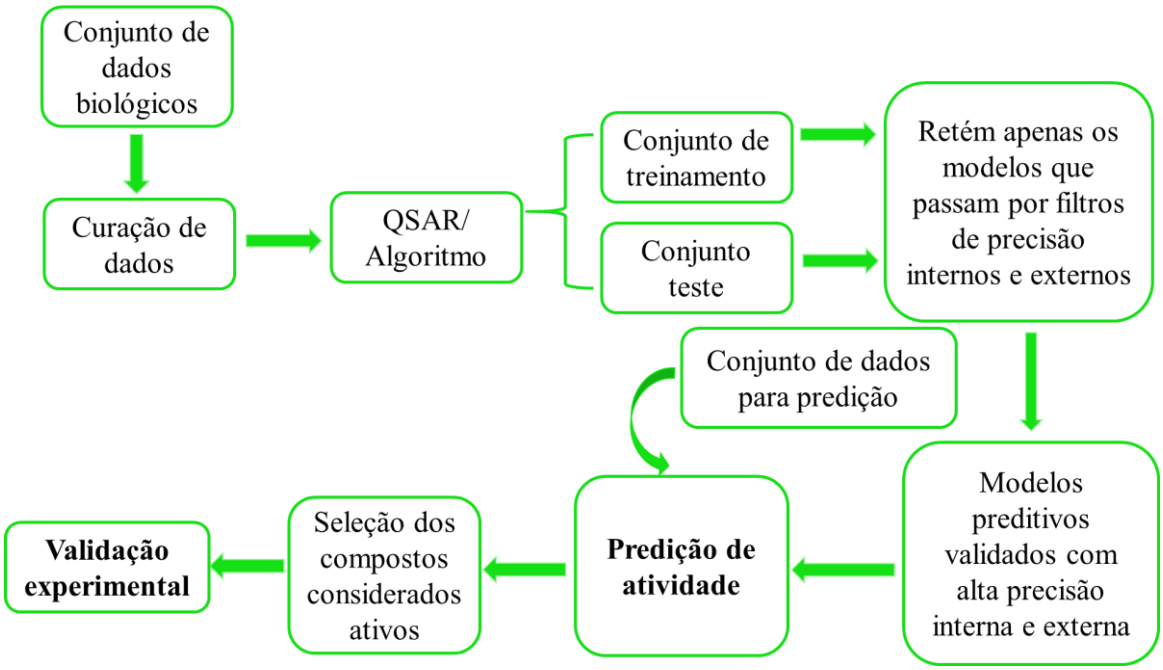
O QSAR teve origem em 1960 com os estudos de Hansch e Fujita, a partir de conhecimentos das propriedades moleculares de compostos orgânicos [73]. Um estudo publicado por Hammett em 1930 [77], sobre a relação linear entre a variação dos grupos substituintes e a propriedade biológica estudada contribuiu para que Hansch e Fujita consolidassem os estudos de QSAR. No estudo realizado por eles, a atividade biológica poderia ser linearmente correlacionada com diferentes parâmetros físico-químicos, relacionados a efeitos hidrofóbicos, estéricos e eletrônicos [78]. Na mesma época, Free e Wilson desenvolveram uma abordagem para descrever a atividade biológica de compostos através de equações estabelecidas a partir da análise de séries congêneres [79]. Em 1970, Kubinyi refinou o modelo de Hansch, desenvolvendo um modelo que descrevia a dependência não linear da atividade biológica sobre o caráter hidrofóbico [50].

A figura 4 mostra uma série de etapas que fazem parte do QSAR.

2.2.2.1 Curadoria, preparo e padronização de dados químicos

Com o recente advento de tecnologias de alto rendimento para síntese de compostos, triagem biológica e depósito de diversas estruturas em banco de dados de estruturas químicas, houve a necessidade de selecionar e organizar bem essa grande quantidade de informações [80].

Figura 4. Fluxo de trabalho de modelagem QSAR preditiva.



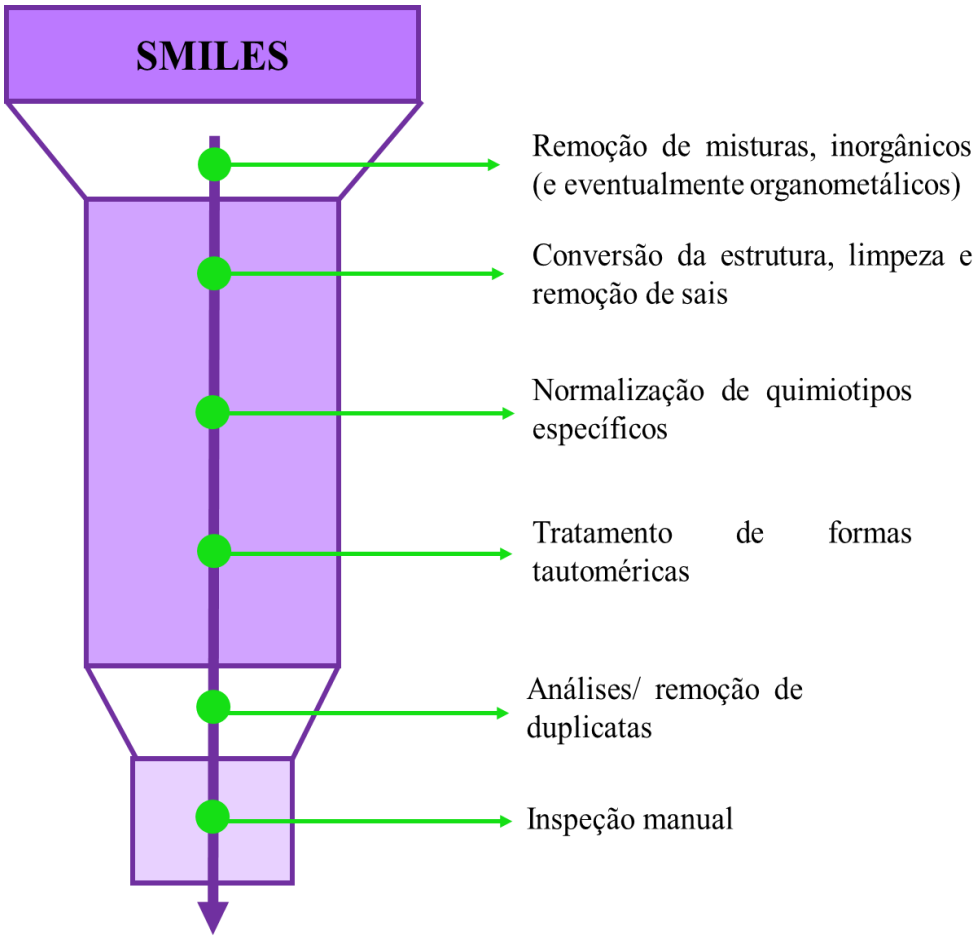
Adaptado de Trospha 2010 [80].

Para a quimioinformática é fundamental a exatidão dos dados de entrada gerados pelos experimentos e disponíveis em vários conjuntos de dados, pois a inconsistência desses dados pode gerar falsos resultados [81].

A curadoria de dados químicos é um dos procedimentos mais relevantes desenvolvidos para aumentar a confiabilidade e o poder preditivo dos modelos em QSAR. Esse procedimento segue um protocolo que consiste na identificação e correção de erros nas estruturas químicas [50]. Segundo Cherkasov, *et al*. (2014) [73], inicialmente são removidas as misturas, por exemplo; compostos orgânicos com compostos inorgânicos, compostos isômeros ou compostos impuros. Na segunda etapa, os sais são removidos e os compostos convertidos. Em seguida, é realizado a normalização de quimiotipos específicos, que consiste na correção de compostos que possuem padrões estruturais diferentes no mesmo conjunto de dados. Na quarta etapa, são removidas as duplicatas e padronizados os dados experimentais dos conjuntos de dados. A análise das duplicatas é importante pois permite avaliar a qualidade dos dados experimentais e remover estruturas químicas de registros duplicados com dados experimentais contraditórios, que afetam a qualidade dos modelos. Por fim, é realizado uma verificação manual, com a finalidade de remover erros não identificados nas etapas anteriores. Consiste basicamente em (i) verificação rápida do nome IUPAC do composto (se disponível) e de sua estrutura; (ii) a normalização dos vínculos é incompleta - erros comuns estão relacionados à presença de diferentes representações dos mesmos grupos funcionais; (iii) algumas duplicatas ainda podem estar presentes apesar do uso de software automatizado para removê-las, por exemplo, alguns tautômeros ainda podem ser encontrados; e (iv) outros erros - cargas erradas, presença de hidrogênios explícitos em uma estrutura empobrecida de hidrogênio, ligações incorretas, etc. Os mesmos autores desenvolveram um modelo de fluxo de trabalho para curadoria de dados que pode ser visualizado na figura 5.

Cada laboratório de quiminformática pode ter seu próprio protocolo para preparar e curar um conjunto de dados de compostos antes da modelagem. No entanto, para um melhor desempenho é necessário observar que algumas informações de bancos de dados usados para modelagem podem ser errônea, o que provavelmente reduzirá a qualidade dos modelos derivados. Será necessário corrigir erros estruturais sempre que possível (às vezes à custa da remoção de ou registros de dados confusos) [81].

Figura 5. Fluxo de trabalho geral de curadoria de conjunto de dados.



Adapatado de Cherkasov, *et al.* (2014) [73].

Após a padronização das moléculas, uma análise dos dados com os valores de atividade biológica de cada composto é feita para remover as estruturas *outliers*. Essa etapa estima a viabilidade de obtenção de modelos preditivos para um determinado conjunto de dados e serve como indicadores adicionais de qualidade dos dados [50]. Por fim, todas as moléculas com seus respectivos códigos SMILES são gerados e usados como dados de entrada em softwares usados para padronizar estruturas químicas, através da adicioção de hidrogênios, aromatização, conversão em três dimensões e, finalmente, salvo em formato indicado pelo usuário.

2.2.2.2 Geração de descritores moleculares

Para a realização do QSAR, é necessário que a série de compostos seja submetida a programas que calculam descritores moleculares de cada estrutura química. O descritor molecular é o resultado final de um procedimento lógico e matemático que codifica a informação química numa representação simbólica de um número útil ou um resultado de algum experimento padronizado [76].

Os descritores moleculares consistem em vários níveis de representação da estrutura química. Esses níveis variam de fórmula molecular unidimensional (1D) à fórmula estrutural bidimensional (2D), a níveis tridimensionais, dependentes de conformação (3D) e até níveis mais elevados, levando em consideração a orientação mútua e dependente do tempo de dinâmica das moléculas (4D) [73]. Outros níveis de representação foram desenvolvidos, como descritores 5D que foram propostos por Vedani e Dobler [82], e permitem a representação das moléculas do ligante por um conjunto de conformações, orientações e estados de protonação (a quarta dimensão), além de permitir simultaneamente até seis diferentes ajustes induzidos locais. Para estimar as energias livres de ligação ao ligante, leva em consideração fenômenos de solvatação, tensão interna e mudanças na entropia durante a ligação ao receptor. O mesmo grupo de pesquisadores desenvolveram em seguida os descritores 6D, que consideram vários modelos de solvatação simultâneos [83]. Também consiste num conjunto de parâmetros capaz de descrever uma molécula em termos quantitativos. Dependendo da representação molecular selecionada, é possível descrever uma molécula por meio de descritores experimentais ou teóricos. Descritores experimentais são todas as medições experimentais, como coeficiente de partição octanol-água, refratividade molar, polarizabilidade e, em geral, qualquer propriedade físico-química obtida aplicando um procedimento experimental bem especificado. Enquanto que os descritores moleculares teóricos são aqueles descritores obtidos por um algoritmo quimioinformático aplicado a uma representação molecular [76].

Na tabela 2, são apresentados os principais tipos de descritores moleculares relatados por Cherkasov *et al* (2013) e [73] Consonni e Todeschini (2017) [76].

Tabela 2. Descrição dos principais tipos de descritores moleculares.

Descritor molecular	Descrição
Descritores 1D	Abrangem todos aqueles descritores que levam em consideração apenas a composição química e não codificam informações sobre a topologia geral e geometria. Exemplo: o peso molecular, o número de átomos e ligações, frequência de ocorrência absoluta e relativa do átomo específico, tipos de ligação e descritores de anel

	obtidos considerando apenas a composição da estrutura molecular.
Descritores 2D	São descritores topológicos que codificam informações diferentes sobre os pares de vértices, conectividade (matriz de adjacência), distâncias topológicas (matriz de distância) e somas dos pesos dos átomos ao longo dos caminhos de conexão (matrizes ponderadas). Também codificam informações químicas sobre os vértices (matrizes aumentadas); além das propriedades atômicas mais comuns, como grau do vértice, soma da distância do vértice, excentricidade do átomo, são freqüentemente encontrados como as ponderações atômicas.
Descritores 3D	São aqueles descritores moleculares calculados levando em em consideração a estrutura tridimensional da molécula. São descritores geométricos derivados de um representação gráfica tridimensional da molécula, levando em consideração não apenas as posições dos átomos, mas também as conexões entre eles. Os descritores geométricos têm um conteúdo de informação mais alto do que os descritores bidimensionais, mas devem ser tratados com cuidado, pois seus valores variam modificando a conformação da molécula.

Na tabela 3 estão listados alguns programas e servidores web que calculam descritores moleculares.

Tabela 3. Lista de alguns softwares para cálculos de descritores moleculares.

Programa/servidor web	Descrição	Ref.
PaDEL-Descriptor	É um software para calcular descritores moleculares e impressões digitais. O software atualmente calcula 797 descritores (663 1D, descritores 2D e 134 descritores 3D) e 10 tipos de impressões digitais.	[84]
Dragon 7.0	Calcula 5.270 descritores moleculares, abrangendo a maioria das várias abordagens teóricas. A lista de descritores inclui os tipos de átomos mais simples, grupos funcionais e contagens de fragmentos, descritores topológicos e geométricos, descritores tridimensionais, mas também propriedades como logP e regra de Lipinski.	[85]
Volsurf	Aborda descritores que quantificam as interações estéricas, hidrofóbicas e de ligações de hidrogênio entre os compostos do modelo e diferentes ambientes. Além disso, tem sido amplamente usado para modelar propriedades farmacocinéticas, por exemplo,	[86]

	permeabilidade passiva através do trato gastrointestinal ou através da barreira hematoencefálica. Calcula 128 descritores.	
MOLGEN-QSPR	Atualmente fornece 705 descritores de vários tipos: • descritores aritméticos (usando informações codificadas na fórmula molecular do composto), • descritores topológicos (usando informações codificadas na constituição do composto), • descritores geométricos (usando informações 3D codificadas na configuração e conformação do composto), • índices eletrotopológicos, • índices gerais, • Índices de Crippen.	[87]
Mordred	O Mordred calcula 1.800 descritores e pode ser facilmente instalado e usado na interface de linha de comando, como um aplicativo da web ou como um pacote Python de alta flexibilidade em todas as principais plataformas (Windows, Linux e macOS). É pelo menos duas vezes mais rápido que o PaDEL-Descriptor e pode calcular descritores para moléculas grandes, o que não pode ser realizado por outro software.	[88]
PRIMoRDiA	Consiste num software construído para calcular descritores de reatividade de grandes biossistemas, empregando um tratamento eficiente e preciso produzidos por pacotes de química quântica.	[89]

2.2.2.3 Método e algoritmos em QSAR

O aprendizado de máquina (do inglês "*Machine learning"* - ML) é um método de inteligência artificial bastante usado em QSAR e consiste em algoritmos e modelos estatísticos que os sistemas de computadores usam para realizar uma tarefa específica sem serem programados explicitamente [90].

A enorme e crescente quantidade de dados e poder computacional têm contribuído para o aperfeiçoamento e desenvolvimento de diversos algoritmos no aprendizado de máquina. O ML geralmente fornece sistemas com a capacidade de aprender e aprimorar a partir da experiência automatizada sem ser especificamente programado e é geralmente referido como a tecnologia mais recente e mais popular na quarta revolução industrial [91]. No entanto, a ideia de um computador aprender algum conceito abstrato a partir de dados e aplicá-los à situações invisíveis não são novas e existem desde 1950 [92]. Esse método tem

sido usado em biologia molecular, descoberta de drogas, desenvolvimento de drogas e farmacologia clínica.

O *Machine learning* (ML) lida com conjunto de dados, composto por vários pontos de dados (também chamados de amostras), onde cada ponto de dados representa uma entidade que se deseja analisar. Cada ponto de dado pode ser representado, por exemplo; por um paciente ou uma amostra de tecido canceroso, ou um composto. Para compilar o conjunto de dados, é necessário obter um número de recursos, que consiste em propriedades que descrevem os pontos de dados. Esses recursos podem ser categóricos (valores predefinidos como masculino e feminino), ordinal (predefinido valores que têm uma ordem intrínseca a eles, como um estágio de doença), ou numérico (por exemplo, valores reais) [92].

Após a obtenção do conjunto de dados, os recursos derivados dos dados são transformados em log, produtos e proporções de recursos, ou combinações mais avançadas. Essa etapa denomina-se pré-processamento e pode ter um efeito profundo sobre o desempenho do modelo [92]. Em geral, a eficácia e a eficiência de uma solução de aprendizado de máquina dependem da natureza e das características dos dados e do desempenho dos algoritmos de aprendizado [91].

Os algoritmos de aprendizagem podem ser categorizados em dois tipos principais, aprendizagem não supervisionada e supervisionada. Não existe um único tipo de algoritmo que sirva para todos ou um que seja o melhor para resolver um problema. O tipo de algoritmo empregado depende do tipo de problema que se deseja resolver, do número de variáveis, do tipo de modelo que melhor se adequaria a ele e assim por diante [90].

Aprendizagem não supervisionada

Métodos de aprendizagem de máquina não supervisionados são particularmente úteis em tarefas de descrição e analisa conjuntos de dados não rotulados sem a necessidade de interferência humana, ou seja, um processo orientado a dados [91], [93]. A finalidade do aprendizado não supervisionado é identificar dimensões, tendências, componentes, *clusters* ou trajetórias subjacentes dentro de uma estrutura de dados. No aprendizado não supervisionado, não há classes ou números para prever. Aqui, é tarefa do programa reconhecer padrões de forma independente nos dados [94]. A aprendizagem não supervisionada pode ser aplicada; por exemplo, para identificar dimensões e trajetórias de saúde mental subjacente ou não

observadas e para determinar a melhor forma de categorizar as dimensões em subtipos (por exemplo, grupos de diagnóstico) [93].

Aprendizagem supervisionada

Segundo Uddin *et al* (2019) [95], o aprendizado de máquina supervisionado usa um conjunto de dados de treinamento rotulado primeiro para treinar o algoritmo subjacente. Este algoritmo treinado é então alimentado no conjunto de dados de teste não rotulado para categorizá-los em grupos semelhantes. O conjunto de dados de entrada é dividido em conjunto de dados de treinamento e teste. Os algoritmos de aprendizagem supervisionada se adaptam bem a dois tipos de problemas: problemas de classificação e problemas de regressão. Em problemas de classificação, a variável de saída subjacente é discreta e é categorizada em diferentes grupos, como 'vermelho' ou 'preto', ou pode ser 'diabético' e 'não diabético'. Em problemas de regressão, a variável de saída correspondente é um valor real como o risco de desenvolver doenças cardiovasculares para um indivíduo. Portanto, o aprendizado de máquina supervisionado infere uma função a partir de dados de treinamento rotulados que consistem em um conjunto de exemplos de treinamento. Os algoritmos de aprendizado de máquina supervisionados são aqueles algoritmos que precisam de assistência externa [91].

No aprendizado supervisionado, um conjunto de variáveis de entrada, como metabólitos no sangue ou níveis de expressão gênica, são usados para prever uma variável de resposta quantitativa, como nível de hormônio, ou qualitativa, como indivíduos saudáveis *versus* indivíduos doentes [96].

A tabela 4 mostra alguns tipos de algoritmos da aprendizagem não supervisionada e aprendizagem supervisionada.

Tabela 4. Resumo de algoritmos de aprendizado de máquina.

Método	Algoritmo	Descrição	Ref.
Aprendizagem não supervisionada	*k-means clustering*	Classifica os dados em grupos k, minimizando as distâncias dentro do grupo para o centroide.	[97]
	Principal component analysis	Usa o procedimento ortogonal para transformar um conjunto de recursos correlacionados em novas variáveis independentes chamadas componentes principais.	[98]

Aprendizagem supervisionada	*Multiple regression analysis*	Consiste num processo estatístico para encontrar relações entre variáveis dependentes e um ou mais independentes variáveis.	[99]
	k-nearest neighbor	Uma aprendizagem baseada em instância onde um objeto é classificado pela regra da maioria entre seu vizinho mais próximo k, onde k é um inteiro.	[100]
	Naive bayes	Uma abordagem probabilística que usa a probabilidade anterior e a regra de Bayes para prever a adesão, assumindo independência de recursos.	[101]
	Decision tree	As instâncias são classificadas por verificação do atributo definido pelo nó, começando na raiz, depois no nó da árvore e, em seguida, desce no galho da árvore correspondente ao valor do atributo.	[91]
	Random forest	Uma técnica de classificação baseada no conjunto de várias árvores de decisão e regras de votação por maioria.	[102]
	Neural network	Um método de aprendizagem baseado em modelo que aprende a partir de dados de entrada com base em camadas de neurônios conectados que consistem em camadas de entrada.	[103]
	Support vector machine	Um método estatístico que mapeia dados em um espaço de alta dimensão para identificar um hiperplano dimensional inferior que maximiza a separação de dados usando um kernel não linear. Isso é obtido maximizando as margens entre hiperplanos conhecidos como vetores de suporte.	[104]

Adapatado de Lo *et al* (2018) [105].

2.2.2.4 *Random Forest* (Floresta Aleatória)

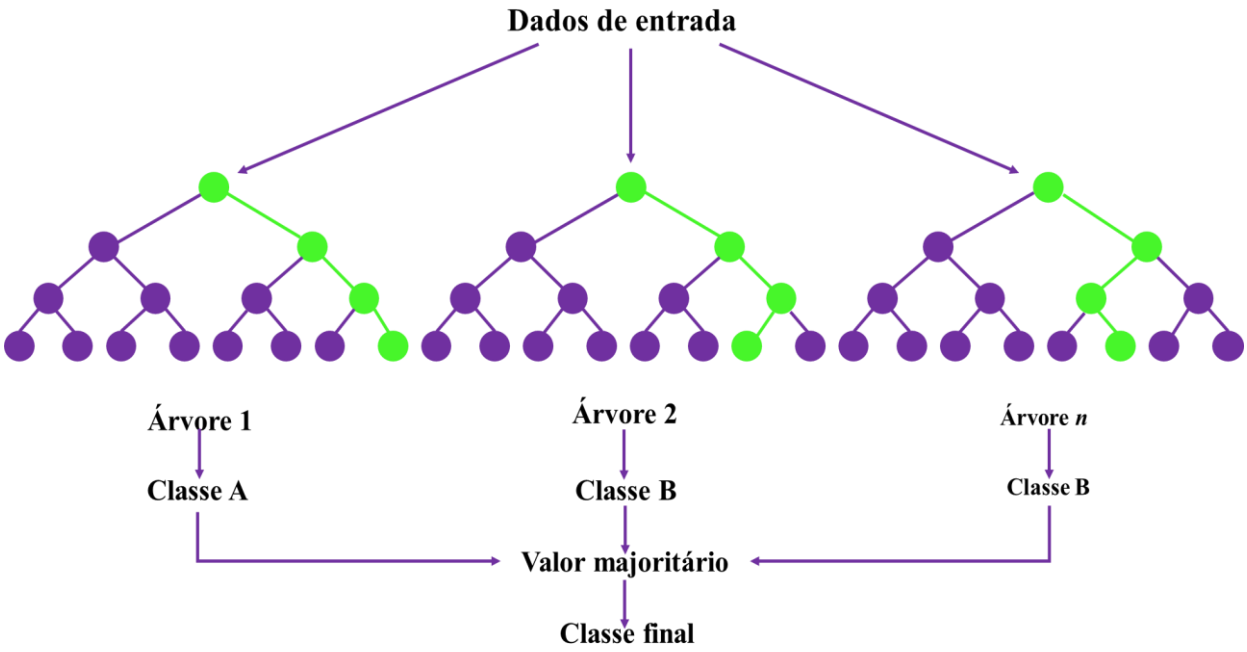
O *Random Fores*t (RF) foi o algoritmo utilizados nesse trabalho e consiste num método supervisionado e um classificador de floresta aleatória bem conhecido, usado no campo de aprendizado de máquina e dados de ciência em várias áreas de aplicação. Esse

algoritmo utiliza várias árvores de decisão como classificadores em paralelo, em diferentes subamostras do conjunto de dados e usa votação por maioria ou médias para o resultado final. Portanto, o RF com várias árvores de decisão é normalmente mais preciso do que um modelo baseado em árvore de decisão única e é adaptável a problemas de classificação e regressão e ajusta bem para valores categóricos e contínuos [91].

O *Random Forest* é uma ferramenta poderosa para modelagem QSAR, pois oferece desempenho de previsão entre os melhores, parece ter um bom desempenho, mesmo sem ajuste de parâmetro e seleção de descritor. Além disso, preserva a maioria dos recursos atraentes das árvores de decisão, como a capacidade de lidar com um grande número de diferentes tipos de descritores simultaneamente, lidar com descritores redundantes/ irrelevantes, incorporar interações e múltiplos mecanismos de ação e capacidade de lidar com classificação e regressão [106].

Segundo Svertnik *et al* (2003) [106], o *Random Forest* é um conjunto de árvores, onde um conjunto de *n* moléculas para treinamento, $D = \{(X_1, Y_1), ..., (X_n, Y_n)\}$; X_i, $_i = 1, ..., n$, é um vetor de descritores e Y_i é o rótulo de classe (por exemplo, ativo/inativo) ou atividade de interesse (por exemplo, $-\log IC_{50}$). O algoritmo de treinamento procede da seguinte maneira: (1) A partir dos dados de treinamento de *n* moléculas, um *bootstrap* da amostra é desenhado (ou seja, amostra aleatória, com substituição, *n* moléculas). (2) Para cada amostra de *bootstrap*, cresce uma árvore com a seguinte modificação: em cada nó, a melhor divisão entre um subconjunto de descritores é selecionada aleatoriamente ao invés de todos os descritores. A árvore cresce até o tamanho máximo (ou seja, até que nenhuma outra divisão seja possível) e não podada de volta. (3) As etapas acima são repetidas até um suficientemente grande número de árvores serem cultivadas (Figura 6). Quando a melhor divisão em cada nó é selecionada entre todos os descritores, as árvores agregadas produzem uma previsão final. Para problemas de classificação, o resultado final é a classe prevista pela maioria das árvores. Na regressão, é a média das previsões de árvores individuais.

Figura 6. Representação do processamento realizado pelo algoritmo *Random Forest*.



Adaptado de https://www.analyticsvidhya.com/blog/2020/12/lets-open-the-black-box-of-random-forests/ Acessado: 09/08/2021.

2.2.2.5 Avaliação de modelos e desempenho

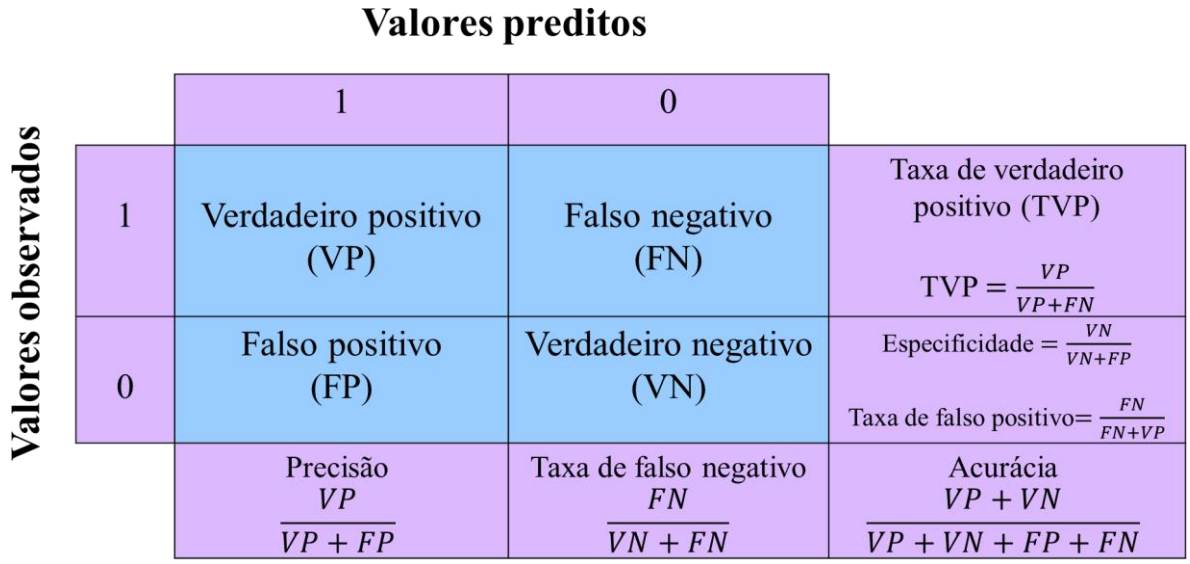
A aplicação de modelos QSAR depende da significância estatística e capacidade preditiva dos modelos [107]. De acordo com Badillo *et al*, 2020 [92], antes da avaliação de um modelo, é importante saber que o princípio geral da seleção do modelo é: quando há dados suficientes, esses são separados em três subconjuntos – conjuntos de treinamento, teste e validação. O conjunto de treinamento é usado para construir diferentes modelos, enquanto o conjunto de validação é subsequentemente usado para escolher o algoritmo e selecionar os hiperparâmetros, se necessário. Por fim, o modelo com melhor desempenho no conjunto de validação é selecionado. Enquanto o conjunto de teste permite avaliar o erro de generalização, que é o erro de previsão sobre um conjunto de dados de teste que não foi usado durante o treinamento.

Matriz de confusão

A matriz de confusão é uma ótima alternativa para verificar ou medir o rendimento do modelo, permitindo a visualização da distribuição de erro de contingência cometido por um

classificador [108]. Considerando um modelo classificador de duas instâncias positivo (P) e negativo (N), as escolhas são estruturadas para predizer a ocorrência ou não de um evento (Figura 7). Assim, VP (verdadeiros positivos): são instâncias corretamente reconhecidas pelo sistema. FN (falsos negativos): são instâncias que são positivos e que o sistema diz que não. FP (falsos positivos): são instâncias que são negativas, mas o sistema diz que não. VN (negativos verdadeiros): são instâncias que são negativas e reconhecidos corretamente assim sendo [108]. A partir dessas informações, é possível calacular a taxa de precisão, que correspondente à razão entre os valores positivos previstos corretamente e o número total de valores positivos previstos; a taxa positiva verdadeira (TPV), que correspondente à proporção de valores positivos previstos corretamente para o número total de valores positivos no conjunto de dados; a taxa de falsos positivos (TFP), que corresponde à proporção de valores negativos previstos incorretamente e a acurácia, que correspondente ao número de valores preditos corretamente divididos pelo número total de valores preditos [92]. Para ser considerado um bom modelo, a precisão, a acurácia e a especificidade devem apresentar resultados acima de 0.6 [50].

Figura 7. Matriz de confusão. A matriz de confusão indica o quão bem-sucedido o algoritmo foi em prever valores em um problema de classificação binária, assumindo valores 0 (chamado de "negativo") ou 1 (chamado de "positivo"), avaliando os valores predito *versus* real.



Adaptado de Badillo *et al*, 2020 [92].

Domínio de aplicabilidade

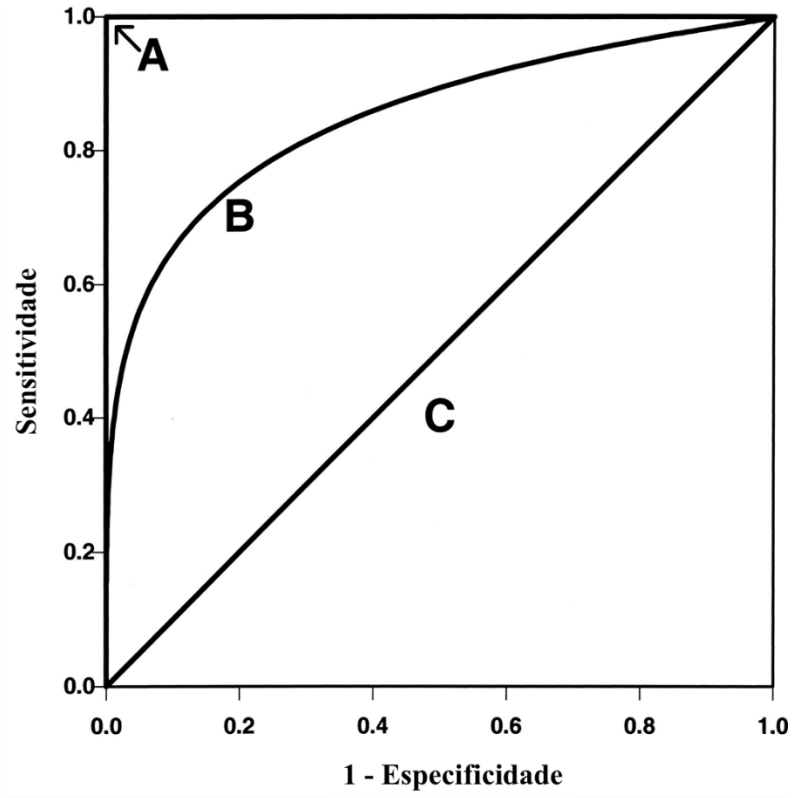
Para que a resposta de previsão de um modelo de QSAR seja confiável, é necessário que o composto que está sendo previsto esteja dentro do domínio de aplicabilidade do modelo. O domínio de aplicabilidade é uma região teórica do espaço químico, definida pelos descritores do modelo e a resposta modelada e, portanto, pela natureza das moléculas do conjunto de treinamento [107]. Portanto, se um composto for altamente diferente de todos os compostos do conjunto de modelagem, é improvável que seja realizada uma previsão confiável de sua atividade [76].

Curva ROC

A curva ROC do inglês "*Receiver operating characteristic*", é um dos classificadores mais amplamente utilizados e consiste numa representação gráfica do desempenho de um modelo e plotagem do dados quantitativos da sensibilidade (proporção de verdadeiros positivos) contra a especificidade (proporção de falsos positivos) para diferentes valores de teste [109] (Figura 8).

A curva ROC também pode ser usada para indicar a dimensão de um efeito. O indicador mais utilizado é a área sob a curva (do inglês "*area under the curve*" - AUC), que é o resultado da integração de todos os pontos ao longo do caminho da curva. A AUC estima o comportamento e precisão do conjunto porque permite calcular a probabilidade de classificar corretamente um sujeito aleatório, na qual um AUC de 0,7 indica uma probabilidade de 70% de classificação correta [110].

Figura 8. Três curvas ROC hipotéticas que representam a precisão diagnóstica do padrão perfeito (linhas A; AUC = 1) nos eixos superior e esquerdo do quadrado da unidade, uma curva ROC típica (curva B; AUC = 0,85) e uma linha diagonal correspondente a chance aleatória (linha C; AUC = 0,5).



Fonte: Polo *et al*, (2020) [110].

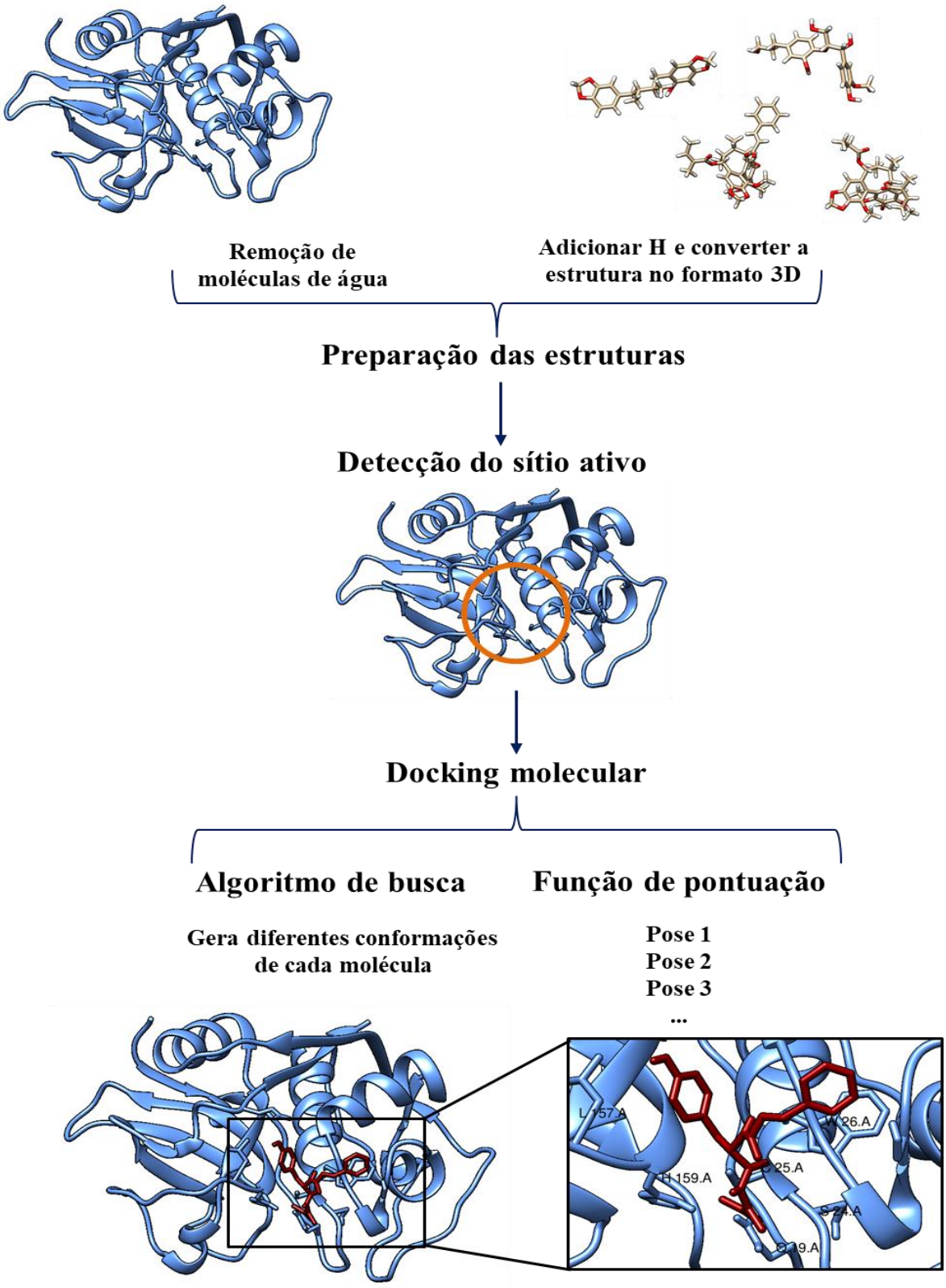
2.2.3 Docking molecular

O docking molecular ou ancoragem é uma ferramenta bastante usada na triagem virtual. Esta técnica consiste em posicionar um ligante dentro do sítio ativo da proteína e encontrar a melhor orientação do ligante que irá formar um complexo com o mínimo de energia global (Figura 9). Os resultados são gerados a partir de pontuações baseadas nos algoritmos a partir de várias combinações possíveis [51], [111], [112].

A previsão da ligação de moléculas à proteínas é de suma importância para o desenvolvimento de novas drogas [113]. Assim, a triagem virtual baseada na estrutura é relevante para o desenvolvimento de agentes terapêuticos e para o entendimento de mecanismos biológicos. Além disso, o uso do docking na triagem virtual contribui para a compreensão da base molecular de uma doença e utiliza o conhecimento da estrutura tridimensional do alvo biológico no processo. A Cristalografia de alto rendimento e métodos de Ressonância Magnética Nuclear (RMN) contribuíram para a aquisição das estruturas

atômicas de proteínas e complexos de proteína-ligante que possibilitaram a obtenção de inibidores moleculares com potenciais usos terapêuticos [114].

Figura 9. Fluxograma de triagem virtual baseada na estrutura do alvo através do docking molecular.



O docking permite analisar interações complexas que auxiliam na predição do mecanismo de ação de moléculas através de leis da termodinâmica. Forças motrizes que ditam a associação entre proteínas e ligantes são um resultado de várias interações e ligações entre proteínas e ligantes. A estabilidade do complexo proteína-ligante é determinada pela magnitude do delta G negativo, que define a afinidade de ligação de qualquer ligante à sua proteína:

$$\Delta G = \Delta H - T\Delta S \qquad (1)$$

onde, ΔG é a energia livre de ligação, ΔH é a entalpia, T é a temperatura em Kelvin e ΔS a entropia [115].

2.2.3.1 Algoritmos

Os algoritmos de docking devem ser capazes de reproduzir o modo de ligação experimental e a função de pontuação deve classificá-la como a mais alta entre todas as conformações geradas [116]. O algoritmo analisa e gera a pose do ligante no local de ligação de um alvo, levando em consideração o graus de liberdade roto-translacionais e internos do ligante [117].

Segundo Torres *et al*. (2019) [118], os algoritmos de docking podem ser classificados como sistemáticos, estocásticos ou determinísticos. Algoritmos sistemáticos exploram o grau de liberdade de cada ligante de forma incremental (Exemplos: Construção incremental e conjuntos conformacionais). Os algoritmos de busca estocástica realizam mudanças aleatórias nos graus de liberdade do ligante, não garantindo convergência para a melhor solução (Exemplos: Monte Carlo, Algoritmos Evolucionários, Pesquisa Tabu e Otimização de enxame). Na busca determinística, a orientação e conformação do ligante em cada interação é determinado pelo estado anterior, e o novo estado tem valor de energia igual ou inferior ao anterior. Neste, envolve minimização de energia e dinâmica molecular. A tabela 5 mostra os tipos de algoritmos existentes conforme relatado por Saikia e Bordoloi [119] e Torres *et al*. [118].

Tabela 5. Algoritmos empregados em softwares de docking.

Algoritmo	Descrição	Programas
	Métodos de construção incremental (do inglês "*Incremental construction*" – IC) usam os ligantes	

Construção incremental	divididos em vários fragmentos quebrando suas ligações rotativas e, em seguida, um desses fragmentos é selecionado para encaixar no site ativo primeiro. Os fragmentos restantes são adicionados incrementalmente.	FlexX [120].
Conjunto conformacional	Esse tipo de algoritmo inclui construção incremental e busca aleatória por conformação utilizando as interações de Coulomb e Lennard-Jones. O aumento do tamanho do conjunto reduz o desempenho dos conjuntos de melhor performance.	DOCK 4.0 [121], FLOG [122].
Pesquisa exaustiva	Todas as translações e rotações possíveis de cada conformador no sítio ativo são enumeradas. Além disso, cada conformador é independente da função de pontuação (ou seja, duas funções de pontuação diferentes irão pontuar o mesmo conjunto exato de poses ao usar a pesquisa exaustiva).	FRED [123], Glide [124] eHiTS [125].
Monte Carlo	Gera rotação de poses, translação ou rotação de corpo rígido, permitindo que o ligante cruze barreiras de energia potencial, não facilmente alcançado por métodos de Dinâmica Molecular (DM).	ICM [126].
Algoritmo genético	Usa o princípio da evolução Darwiana e considera cada conformação de um ligante como um cromossomo constituído por genes de valor real que representam graus de liberdade de translação, orientação e conformação do ligante. Os indivíduos são representados pelos compostos, que são avaliados por uma função de pontuação. Os genes de um indivíduo passam por mutações, gerando uma nova prole que será testada.	AutoDock [127], AutoDock Vina [113], Gold [128] e Dockthor [129].
Evolução diferencial	Combina a otimização de evolução diferencial e um algoritmo de previsão de cavidade. Por exemplo, para cada indivíduo na população, uma prole é criada adicionando uma diferença ponderada das soluções pai, que são selecionadas aleatoriamente da população. Depois disso, a prole substitui o pai, se e somente se for mais adequado. Caso contrário, o pai sobrevive e é passado para a próxima geração. Além disso, esse algoritmo usa a previsão de cavidades durante o processo de pesquisa, permitindo uma identificação rápida e precisa dos modos de ligação (poses).	Molegro Virtual Docker [130].
	Abordagem metaheurística; inspirado por meio de	

Otimização de colônias de formigas	comunicação de desenho de feromônio e seguindo o comportamento de formigas do mundo real. Assim, o algoritmo considera para cada um dos três graus de liberdade translacionais um comprimento de intervalo de 0,1, enquanto para os três graus de liberdade rotacionais e todos os graus de torção de liberdade um intervalo de 1 é tomado. Portanto, cada vetor de feromônio está associado a graus de rotação ou torção, enquanto que o número de entradas dos vetores de feromônio correspondente aos três graus de liberdade translacionais dependente do diâmetro do local de ligação. Uma trilha de feromônio *ij*, então, se refere à desejabilidade de atribuir o valor *j* ao grau de liberdade *i*.	PLANTS [131].
Minimização de energia	Conduz o sistema para o mínimo potencial de energia; usado como um método local de pesquisa usado em docagem de relaxamento molecular, dá uma implementação precisa quando usado como um ferramenta de otimização.	MCSS [132].
Dinâmica molecular	Considera o estado termodinâmico e outras variáveis para medir o movimento dos átomos ao longo do tempo.	CDOCKER [133].

2.2.3.2 Funções de pontuação

A função de pontuação tem como objetivos: classificar e predizer a afinidade de ligação absoluta entre o complexo proteína-ligante [134]. Assim, as funções de pontuação são amplamente empregadas no estudo de docking molecular para avaliar as interações proteína-ligante e atualmente podem ser classificadas em três principais grupos: baseado em campo de força, empírico e baseado em conhecimento [116]. Porém, com os avanços nas pesquisas, funções de pontuação baseados em aprendizado de máquina também tem sido utilizado. A tabela 6 mostra a classificação das funções de pontuação e suas subdivisões como descrito por Li *et al*, 2019 [134].

Tabela 6. Categorias das funções de pontuação do docking.

Classificação	Subdivisão	Descrição	Equação
	Campo de força clássico	Calcula a energia de ligação acumulando a interação Van der Waals e eletrostática entre os pares de átomos de proteína-ligante.	$E_{bind} = E_{vdw} + E_{elec}$

Baseado em física	Modelos de Solvatação	Incorpora a entropia de torção de ligantes e o efeito de solvatação.	$E_{bind} = E_{vdw} + E_{elec} + \Delta G_{solv}$
	Mecânica Quântica	Aborda interações covalentes, polarização e transferência de carga no docking. A Mecânica Quântica (MQ) acoplada à Mecânica Molecular melhora a precisão preditiva.	$E_{bin} = E_{QM/MM} + \Delta G_{solv}$
Empírico		Emprega um conjunto de dados de treinamento com afinidades de ligação conhecida para otimizar as estruturas de proteína-ligante; selecionando sistematicamente por meio de fatores energéticos importantes para a ligação proteína-ligante, como ligações de hidrogênio, efeitos hidrofóbicos, colisões estéricas, entre outros.	$E_{bind} = E_{vdw} + E_{elec} + \Delta G_{solv}$
Baseado em conhecimento		Geram os potenciais de pares emparelhados de estruturas tridimensionais de um grande conjunto de complexos proteína-ligante baseados no princípio da estatística inversa de Boltzmann.	$E_{bin} = \sum^{L}_{i=1} \sum^{R}_{j=1} w_{ij}®$
Baseado em Aprendizado de Máquina	Máquina de Vetores de Suporte	Emprega uma variedade de algoritmos de aprendizagem de máquina e depende de um conjunto de dados de treinamento.	Discrimina poses ativas de não-ativas.
	Floresta aleatória		Utiliza a contagem de pares de proteína-ligante emparelhados em alguma distância na predição de afinidade de ligação por meio de descritores.
	Rede neural artificial		Considera mais as características de ligação e a saída pode prever a afinidade de ligação por meio de descritores.
	Aprendizagem Profunda		Contribui na previsão de afinidades de ligação proteína-ligante e explora os ligantes ativos de não-ativos.

A **tabela 7** mostra alguns programas de docking e seus respectivos algoritmos e tipo de função de pontuação.

Tabela 7. Programas de docking e seus respectivos algoritmos e funções de pontuação.

Programa	Função de pontuação	Algoritmo
Dock	Campo de força	Construção Incremental
AutoDock	Semi-empírico	Algoritmo genético
AutoDock Vina	Empírico	Algoritmo genético
FlexX	Baseado no conhecimento	Construção Incremental
Gold	Empírico	Algoritmo genético
Glide (Schrodinger)	Emípirico	Pesquisa exaustiva
Fred	Pontuação gaussiana/Empírico	Conjuntos conformacionais
Molegro Virtual Docker	Semi-empírico	MolDock SE/ Evolução diferencial
Plants	Empírico	Otimização de enxames
Fred	Empírico	Complementaridade de forma
ICM	Baseado em física	Monte Carlo
Dockthor	Baseado em física + empírico	Algoritmo genético

Fote: Torres *et al* (2019) [118].

2.2.3.3 Docking consenso e *scoring* consenso

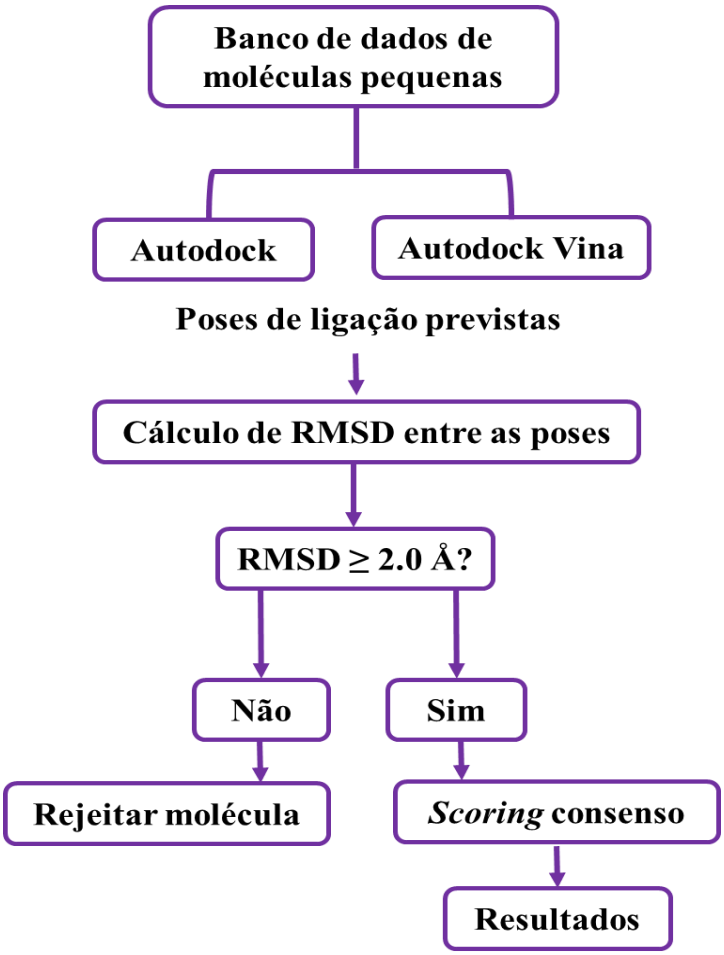
Com o desenvolvimento contínuo de novas funções de pontuação e programas de docking, o uso de estratégias de docking consenso que combinam duas ou mais funções de pontuação ou diferentes programas, tem se tornado cada vez mais comum. Isso é especialmente interessante, porque as várias funções de pontuação disponíveis têm um desempenho diferente em todo o espectro de interações potenciais e, presumivelmente, em uma combinação ideal, as deficiências de uma função específica podem ser compensadas pelas outras [118].

Estudos relataram que a pontuação consenso [135], [136], [137], [138], [139], que combina várias funções de pontuação de um ou mais programas de docking para a definição de afinidade de ligação, pode obviamente melhorar as taxas de acertos em estudos de docking [137].

Existem vários métodos de pontuação consenso que consideram os valores de energia e/ou as poses em suas análises. Um estudo realizado Houston e Walkinshaw (2013) [140], mostrou que esquemas de pontuação consenso, aumentam o poder de classificação e, portanto, as taxas de acerto, combinando informações sobre modos de ligação previstos em vez de afinidades de ligação previstas. O estudo usou dois programas de docking, o Autodock e o Autodock Vina para prever a postura de ligação correta por meio do desvio da raiz quadrada média (RMSD). Os resultados mostraram que de um total de 228 ligantes encaixados em suas respectivas proteínas usando Autodock e Vina separadamente, 122 dos acoplamentos com o Autodock e 141 dos acoplamentos com o Vina foram encontrados para ser posicionado corretamente (RMSD $\leq$ 2,0 Å entre o encaixado e a pose cristalográfica). Um total de 118 ligantes foi encontrado para ser encaixado de forma semelhante nos dois programas (RMSD $\leq$ 2,0 Å), e 97 dessas previsões de modo de ligação estavam corretas (ou seja, o RMSD entre a pose encaixada e a pose cristalográfica foi inferior a 2,0 Å). Usando ambos os programas juntos, uma postura correta é prevista em 82% dos casos. A taxa de sucesso de cada programa sozinho foi de 55% para o Autodock e 64% para o Vina. Assim, combinar os resultados dos dois programas fornece uma maneira de melhorar a confiança nos resultados de encaixe. Um esquema do método pode ser visualizado na figura 10.

Segundo Maia *et al*. (2020) [141], apesar das taxas de acertos em abordagens de docking consenso excederem 60%, a confiabilidade da análise consenso não depende apenas do número de amostras testadas, nem das fórmulas usadas nos métodos de pontuação, mas também do bom desempenho do algoritmo. Isso porque se um estudo usa cinco programas diferentes para obter um pontuação consenso e se apenas um dos algoritmos é responsável pela maior taxa de precisão em comparação com as outras, usando o método de pontuação de consenso, os riscos aumetam, em vez de aumentar a eficiência da técnica. Por isso, os autores recomendam testar e classificar os algoritmos antes de aplicar qualquer estratégia de pontuação consenso.

Figura 10. Fluxo de trabalho de docking consenso baseado no RMSD.



Adaptado de Houston e Walkinshaw (2013) [140].

2.2.4 Simulações de dinâmica molecular (DM)

Uma das principais limitações do docking molecular é considerar a proteína como uma molécula rígida ou parcialmente rígida. Assim, a flexibilidade da proteína continua sendo um desafio para estudos de docking molecular. Por isso, uma estratégia bastante aceita na busca virtual de compostos e suas melhores conformações, é a dinâmica molecular (DM). Essa técnica é frequentemente usada para compreender processos dinâmicos de interações entre proteínas e ligantes, em diferentes escalas de tempo e com detalhes atomísticos a fim de racionalizar fenômenos biológicos [142], [143].

Segundo Hollingsworth e Dror (2018) [144], a DM pode ser usada para responder a várias perguntas. A aplicação mais básica e intuitiva da simulação é avaliar a mobilidade ou flexibilidade de várias regiões de uma biomolécula. Porém, também pode revelar o comportamento de ligantes complexados à uma proteína na presença de moléculas de água e

íons. Frequentemente é utilizada para testar poses de ligação modeladas de ligantes e verificar qual pose é mais estável. Além disso, é possível determinar como um sistema biomolecular responderá a alguma perturbação. Por exemplo: se remover um ligante complexado de uma estrutura proteica determinada experimentalmente e depois simular para ver como a remoção do ligante afeta a conformação da proteína; mutar um ou mais resíduos de aminoácidos na proteína para explicar ou prever a função efeito de uma mutação; para recuperar a estrutura nativa quando a estrutura experimental resolvida diferiu do tipo selvagem; mudar o estado de protonação de um aminoácido ácido ou básico.

Fundamentalmente, a DM é um método *in silico* que estuda o movimento dos átomos e moléculas, através de equações do movimento de Newton sob a representação de um sistema, função energética, conhecido como campo de força. Este envolve a combinação de princípios de física, cálculos de mecânica quântica e informações experimentais [145]. Nas simulações de DM, o campo de força é descrito de forma clássica usando uma função de energia potencial empírica ($V(r)$) [51].

Na equação 2, uma força (f) atua em um átomo i de um sistema num determinado instante que corresponde à sua massa (m) multiplicada por sua aceleração (a). A configuração do sistema neste instante é representada por um vetor r, que descreve a posição de n átomos que interagem no espaço cartesiano (r={x1,y1,z1,x2,y2,z2,..,xN,yN,zN}) [146].

$$f = m.a \tag{2}$$

A Equação 3, é uma representação detalhada do campo de força decorrente da equação 2 simplificada. A aplicação dos três primeiros termos representam interações intramoleculares dos átomos, que descrevem variações na energia potencial como uma função do alongamento, dobra e torções da ligação entre os átomos diretamente envolvidos nas relações de ligação. Esses são representados por somas sobre comprimentos de ligação (l), ângulos (α) e ângulos diédricos (θ), respectivamente. As contribuições de alongamento e flexão da ligação compartilham a mesma forma funcional, pois são ambas descritas por potenciais harmônicos com valores de referência l_0 e α_0 e constantes de força k_l and k_α, respectivamente. Porém, por causa de sua periodicidade intrínseca, os termos de torção são definidos naturalmente por uma série de cossenos de termos para cada ângulo diédrico. Assim, n_{ik} é um parâmetro que descreve a multiplicidade para o quinto termo da série, $\theta_{0,ik}$ é o ângulo de fase correspondente e V_{ik} é a barreira de energia [146]. Esse grupo de contribuições é geralmente referido como os termos "vinculados" do campo de força.

$$V = \sum_{i}^{ligações} \frac{K_{l,y}}{2}\left(l_i - l_{0,i}\right)^2 + \sum_{i}^{ângulos} \frac{K_{a,y}}{2}\left(a_i - a_{0,i}\right)^2 +$$

$$\sum_{i}^{torsões} \{\sum_{k}^{M} \frac{V_{ik}}{2}\left[1 + \cos\left(n_{ik}.\theta - \theta_{0,ik}\right)\right]\} \quad (3)$$

$$+ \sum_{i,j}^{pares} \varepsilon_{ij}[\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6} + \sum_{i,j}^{pares} \frac{q_i q_i}{4\pi\varepsilon_{0\varepsilon}\mathrm{rij}}$$

Embora a DM contribua para a compreensão de diversos mecanismos moleculares, segundo Liu *et al* (2018) [147], o uso da DM na triagem virtual têm uma série de vantagens. Dentre elas:

✓ Pode melhorar o fator de enriquecimento de triagem virtual considerando múltiplas conformações do alvo durante o encaixe e dando uma classificação mais precisa de compostos de sucesso em combinação com cálculos de energia livre de ligação.
✓ Melhora a eficiência da otimização do *lead*, podendo ser usado para obter modos de ligação precisos.
✓ Contribui para descobrir o mecanismo de resistência a drogas.

No tentanto, segundo Childersa e Daggett (2017) [148], nenhuma técnica é sem erro ou limitação, e simulações moleculares não são exceção. A primeira fonte de erro é a imprecisão dos campos de força, que embora tenham sido parametrizados empiricamente, nenhum campo de força pode ser verdadeiramente preciso. No entanto, validações recentes de campo de força de proteínas mostraram que os campos de força contemporâneos podem reproduzir observações experimentais. A segunda principal fonte de erro é a extensão da amostragem conformacional obtida a partir da simulação, ou seja, o tempo de computação e métodos que limitam a flexibilidade e amostragem da proteína. Simulações muito curtas são ruins e muito longas nem sempre são melhores. Isso porque erros numéricos, como erros de arredondamento e erros de truncamento, se propagam à medida que a simulação é estendida como resultado da integração numérica. Porém, algoritmos de integração especializados e outros métodos computacionais para a realização de simulações foram desenvolvidos para mitigar o erro numérico tanto quanto possível.

2.2.4.1 Campo de força

Campos de força se referem a formas funcionais empíricas e conjuntos de parâmetros, para calcular a energia potencial de sistemas biomoleculares [149]. Um campo de força é uma expressão matemática que compreende a forma funcional do potencial de energia, que inclui as possíveis ligações (ligações, ângulos e diédricas) e interação não-ligada (potenciais de van der Waals, potenciais de Coulomb e Lennard-Jones) [150].

Diversos campos de força foram desenvolvidos para descrever biomoléculas em ambientes aquosos; dentre eles destacam-se o AMBER (Construção de modelos assistidos e refinamento de energia), CHARMM (Química na Mecânica Molecular de Harvard), GROMOS96 (Simulação Molecular de Groningen), e OPLS (Potencial otimizado para simulações líquidas) (Tabela 8) [151].

A escolha do modelo de solvente também é essencial e validar o uso de um determinado modelo de água com seus correspondentes campos de força é necessário. Até o momento, existem três modelos de água: o TIP3P (módulos intermoleculares transferíveis de três locais potenciais), o SPC (taxa pontual simples) e o TIP4P, foram desenvolvidos nos últimos anos para reproduzir as propriedades estruturais, dinâmicas e termodinâmicas da água para uma melhor comparação dos experimentos (Tabela 8) [149].

Tabela 8. Campos de força biomoleculares e seu respectivo modelo de água padrão.

Campo de força	Conjunto de parâmetros	Modelo de água	Referência
AMBER	ff99	TIP3P	[152]
	ff99SB	TIP3P	[153]
	ff99SB-ILDN	TIP3P	[154]
	ff99SB-ILDN-phi	TIP4P	[155]
	f99SB-ILDN-NMR	TIP4P	[156]
	ff03	TIP3P	[157]
	ff03ws	TIP4P/2005	[158]
CHARMM	CHARMM22	TIP3P modificado	[159]
	CHARMM22/CMAP	TIP3P modificado	[160]
	CHARMM36	TIP3P modificado	[161]
GROMOS	GROMOS96 (43a1)	SPC	[162]
	GROMOS96 (53a6)	SPC	[163]
	GROMOS96 (54a7)	SPC	[164]
OPLS	OPLS-AA	Nenhum padrão de modelo	[165]

Adaptado de Chong *et al* (2017) [149].

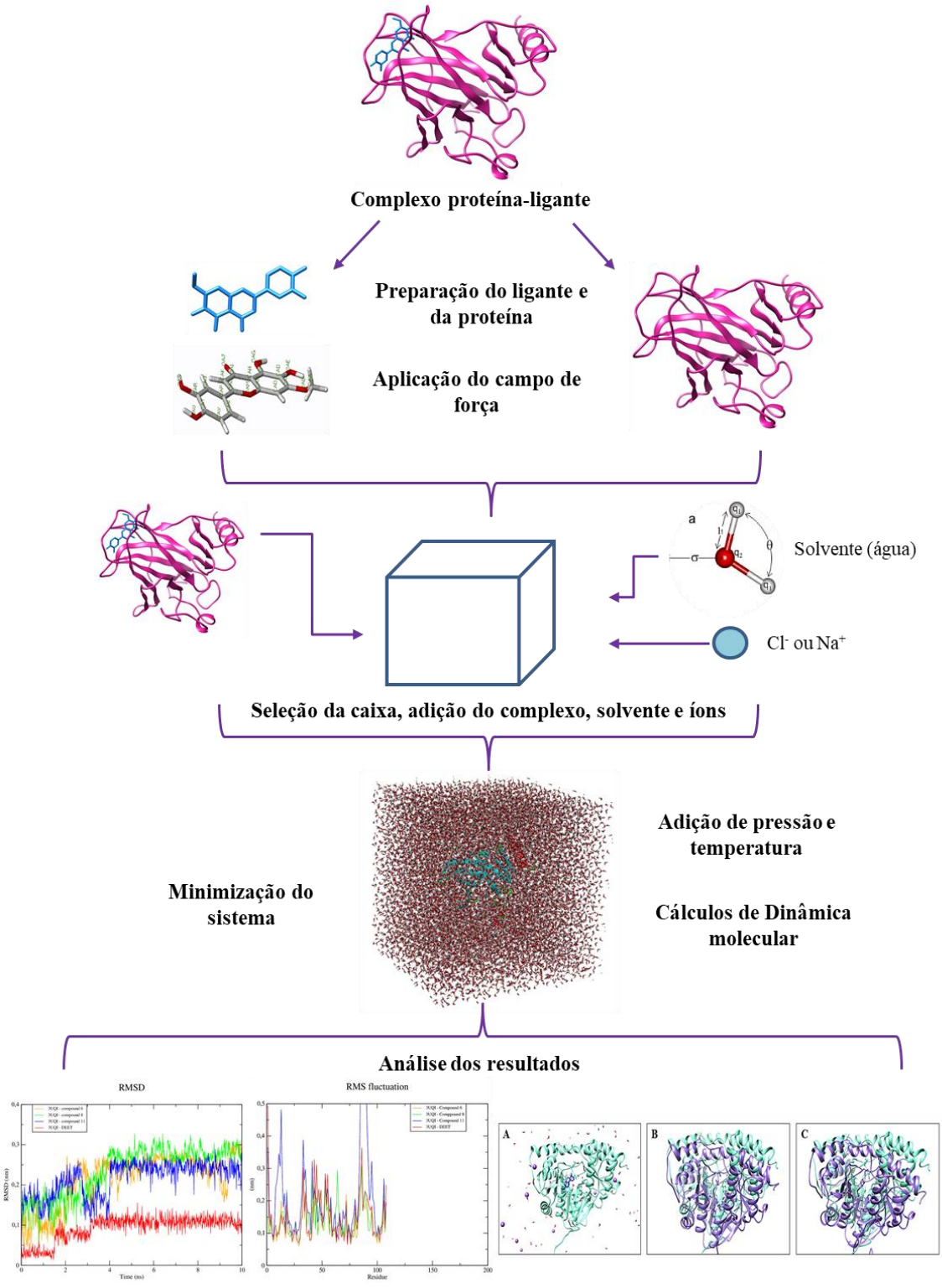
2.2.4.2 GROMACS

A dinâmica molecular gera configurações sucessivas de um dado sistema molecular e geralmente, o produto final é uma trajetória que descreve as posições e velocidades das partículas no sistema ao longo da simulação, pois elas variam com o tempo. Essas simulações avaliam milhões de interações de partículas em diferentes etapas de tempo, que podem e requerem quantidades extraordinárias de hardware computacional e tempo [166]. Atualmente, existem vários programas de DM, dentre eles destacam-se: o GROMACS [166], o AMBER [167], o NAMD [168] e o CHARMM [169].

O GROMACS é um programa utilizado para realizar simulações de dinâmica molecular, desenvolvido pela Universidade de Groningen, na Holanda [170]. O GROMACS é um dos pacotes mais rápidos e mais populares de *software*s disponíveis para DM. Além disso, é extremamente rápido devido à otimizações algorítmicas, operado através de linha de comando, podendo usar arquivos para entrada e saída. Além disso, o programa contém um *script* para converter coordenadas moleculares de um arquivo PDB para gerar topologia e arquivos que o próprio *software* utiliza. A execução da simulação produz arquivos de trajetória descrevendo os movimentos dos átomos ao longo do tempo, que podem ser analisados ou visualizados através de programas gráficos.

Os cálculos de DM envolvem várias etapas, mas a principal delas é a preparação do sistema. Astuti *et al* (2011) [170], descrevem essas etapas e o fluxo de trabalho na figura 11 mostra cada uma delas para DM de complexo proteína-ligante. A preparação do sistema é extremamente importante e inicia-se antes dos cálculos representados pelas equações de Newton. Assim como no docking molecular, a primeira etapa envolve a verificação e preparação do ligante e da proteína. No caso de estruturas baixadas do PDB, é importante verificar se não há átomos ou sequências proteicas ausentes. Caso esteja ausente, é possível modelar usando o programa MODELLER [171]. O formato pdb de ligantes e macromoléculas contém descrições não apenas das coordenadas tridimensionais de seus átomos, mas também a descrição de como esses átomos estão conectados e como eles interagem entre si. Porém, esses arquivos devem ser submetidos a campos de força, gerando uma topologia contendo informações sobre tipos de átomos, cargas, ligações, entre outros.

Figura 11. Etapas de simulações de dinâmica molecular.



No caso do GROMACS, o programa não gera topologia de ligantes. Portanto, devem ser submetidos a campos de força em outros programas ou servidores web. Após a aplicação do campo de força do ligante e da proteína, um ambiente solvatado deverá ser preparado. Para isso, uma caixa cúbica ou dodecaédrica é gerada com as dimensões parametrizadas pelo programa. Milhares de moléculas de água são adicionadas ao sistema; além de conter o complexo, a escolha do tipo de água depende do tipo de campo de força. Depois da adição do solvente no sistema contendo o complexo, o GROMACS calcula a carga líquida do sistema, que deverá ser neutralizada adicionando-se íons. Para remover algumas "imperfeições" que o sistema apresenta, a melhor forma de corrigir é através de uma minimização de energia do sistema. Esse método de otimização essencialmente força um conjunto de átomos a aderir, da melhor forma possível, às definições do campo de força. Em seguida, é adicionado temperatura e pressão através de algoritmos parametrizados pelo GROMACS. Por fim, o sistema está completo para iniciar os cálculos de dinâmica molecular.

2.2.5 Cálculos de energia livre

O reconhecimento molecular é de extrema importância em vários processos bioquímicos e biológicos. Embora o docking combinado com simulações de DM possa fornecer informações importantes sobre a ligação entre proteína-ligante e se essas interações são significativas ou realista, requer uma peça adicional de informação, nomeadamente energia livre de ligação, que é a força motriz para formar este complexo [150]. Portanto, os cálculos de energia livre de ligação são usados para quantificar a afinidade de um ligante ao seu alvo. Na prática, normalmente dá resultados de qualidade, muitas vezes melhor do que o docking molecular [172].

Uma variedade de abordagens computacionais podem ser usadas para estimar as energias livres de ligação, estas incluem: perturbação de energia grátis (do inglês "*free energy perturbation*" - FEP), integração termodinâmica (do inglês "*thermodynamic integration*" - TI), energias de interação linear (do inglês "*linear interaction energies*" - LIE), mecânica molecular Área de superfície de Poisson-Boltzmann (do inglês "*molecular mechanics Poisson−Boltzmann surface area*" - MM-PBSA) e molecular mecânica área de superfície de nascimento generalizada (do inglês "*molecular mechanics Generalized Born surface area*" - MM-GBSA). Porém, destes, o FEP e o TI são os mais rigorosos e os mais caros do ponto de vista computacional. Esses dois métodos envolvem a transformação de uma molécula de um

estado inicial para o estado final ao longo de um caminho reversível, sendo principalmente aplicáveis a pequenas perturbações ou transições estruturais [173].

2.2.5.1 MM-PBSA

O método MM-PBSA permite estimar a energia livre de ligação de complexação de trajetórias de dinâmica molecular e pode ser aplicado em buscas virtuais ou docking para refinar a classificação de poses. Este método aproxima a energia livre de ligação de um complexo por um ciclo termodinâmico no qual a entalpia de solvatação é modelada por um modelo de solvente contínuo e uma função linear de solvente [174]. Nesta técnica, $\Delta G_{ligação}$ é avaliado a partir das energias livres do ligante e da proteína na equação 4 [175]:

$$\Delta G_{ligação} = G_{complexo} - (G_{proteína} + G_{ligante}) \qquad (4)$$

A energia livre para cada entidade individual é dada pela equação 5 onde, *x* simboliza a proteína ou ligante ou o complexo proteína-ligante, E_{MM} é a energia potencial da mecânica molecular em vácuo, e $G_{solvatação}$ é a energia livre de solvatação. Aqui, TS denota a contribuição entrópica para a energia livre em vácuo, onde T e S representam a temperatura e entropia, respectivamente [175].

$$G_x = (E_{MM}) - TS + (G_{solvatação}) \qquad (5)$$

A energia livre de solvatação é a soma da energia livre polar e apolar [173].

$$G_{solvatação} = G_{polar} + G_{não\ polar} \qquad (6)$$

E_{MM} contém as interações ligadas e não ligadas, onde a energia das ligações ligadas que consistem em ligação, ângulo, interações diédricas e impróprias. As interações não ligadas incluem interações eletrostáticas (E_{elec}) e de van interações der Waals (E_{vdw}) [173].

$$E_{MM} = E_{ligada} + E_{não\ ligada} = E_{ligada} + (E_{vdw} + E_{elec}) \qquad (7)$$

Segundo Genheden; Ryde (2015) [172], as simulações usadas para estimar o conjunto energias das equações 6 e 7 que empregam modelos de solventes explícitos. No método MM-PBSA, todas as moléculas de solvente são removidas porque os modelos de solvente implícitos são usados para estimar as energias de solvatação. A energia de solvatação polar representa a interação eletrostática entre o soluto e o solvente contínuo, abrangendo três

termos adicionais necessários para reproduzir experimentalmente as energias de solvatação: cavitação, dispersão e repulsão. A energia de solvatação apolar, considera as partes atraentes e repulsivas das interações geradas pela formação de cavidade e interações de van der Waals entre o soluto e o solvente.

2.2.6 Propriedades de absorção, distribuição, metabolismo, excreção e toxicidade (ADMET)

Inicialmente, num processo de descoberta de drogas, não são considerados algumas propriedades como solubilidade, formação de agregados e estabilidade. Por isso, muitos compostos que chegam aos testes pré-clínicos não conseguem passar dessa fase. Isso porque nos testes pré-clínicos são consideradas propriedades que envolvem absorção, distribuição, metabolismo, excreção e toxicidade (ADMET); que geralmente não são avaliadas nas fases iniciais de descoberta de drogas [51].

Atualmente existem diversas ferramentas computacionais e regras capazes de predizer ou calcular a probabilidade de um composto ter boas propriedades farmacocinéticas e farmacodinâmicas. Assim, faz-se necessário evitar o quanto antes resultados falso-positivos, observando os parâmetros ADMET. Assim, avaliar as propriedades ADMET em compostos candidatos a fármacos no início da descoberta de drogas consiste numa metodologia de filtragem de moléculas relevantes com mais chance de serem aprovadas nos testes pré-clínicos e clínicos. A absorção, distribuição, metabolismo, excreção e toxicidade são baseadas em vários parâmetros e propriedades físico-químicas, farmacocinéticas e farmacodinâmicas que são discutidas nos subtópicos seguintes.

2.2.6.1 Propriedades físico-químicas

Os descritores moleculares físico-químicos são simples como peso molecular (PM), números de ligações rotativas, contagem de tipos de átomos específicos e área de superfície polar topológica (do inglês "*Topologic Polar Surface Area*" - TPSA). Em geral, esses descritores são utilizados para predizer a absorção de uma droga *in vivo*. Valores ideais para cada uma das propriedades físico-químicas citadas configuram um composto com boa absorção, permeabilidade e biodisponibilidade.

Outros descritores que incluem propriedades físico-químicas são o score de biodisponibilidade e a regra de Lipinski. Esses descritores verificam a presença de propriedades de compostos com boas características farmacocinéticas. A biodisponibilidade é um parâmetro composto de solubilidade, permeabilidade e liberação dependente da lipofilicidade [176]. O método que prevê a probabilidade de biodisponibilidade (F) atribui uma pontuação de o composto ter F>10% em ratos. As propriedades físicas relacionadas à permeabilidade ou biodisponibilidade são: peso molecular, log P de octanol-água, log D, área de superfície polar e ligações rotativas [177]. Já a regra de Lipinski prediz a má absorção ou permeação quando há mais de duas violações nos seguintes critérios: não possuir mais de cinco doadores de hidrogênio (H), não possuir mais de dez aceptores de H, o peso molecular ser inferior a 500g/mol e o log P (CLogP) menor que 5 (ou MlogP 4.15) [178].

2.2.6.2 Lipofilicidade

O descritor mais comum para lipofilicidade é o coeficiente de partição entre n-octanol e água (log P). O sistema de n-octanol/água é tradicionalmente empregado em biomedicina e pesquisa farmacêutica. Devido à sua natureza anfifílica, o n-octanol é considerado um bom mimetizador da membrana fosfolipídica [179]. Segundo Arnott e Planey (2012) [176], a lipofilicidade é um parâmetro importante na descoberta e desenho de medicamentos porque constitui a propriedade físico-química mais informativa e bem-sucedida em medicamentos, contribuindo com informações sobre a solubilidade e permeabilidade através de membranas, potência, seletividade, impactando no metabolismo e farmacocinética; e também afetando o perfil farmacodinâmico e toxicológico. Cinco modelos preditivos disponíveis no SwissADME (http://www.swissadme.ch/) aumentam as chances de um modelo preditivo consenso, são eles: XLOGP3, método atomístico que inclui fatores corretivos e biblioteca baseada em conhecimento [180]; WLOGP, método atomístico baseado no sistema fragmentário; MLOGP, método topológico baseado em uma relação linear com 13 descritores moleculares implementados; SILICOS-IT, método híbrido que conta com 27 fragmentos e 7 descritores; e o iLOGP [179], método físico baseado em energias livres de solvatação em n-octanol e água calculado pelo modelo de área de superfície. O log P consenso é a média aritmética dos valores previstos pelos cinco métodos propostos [181].

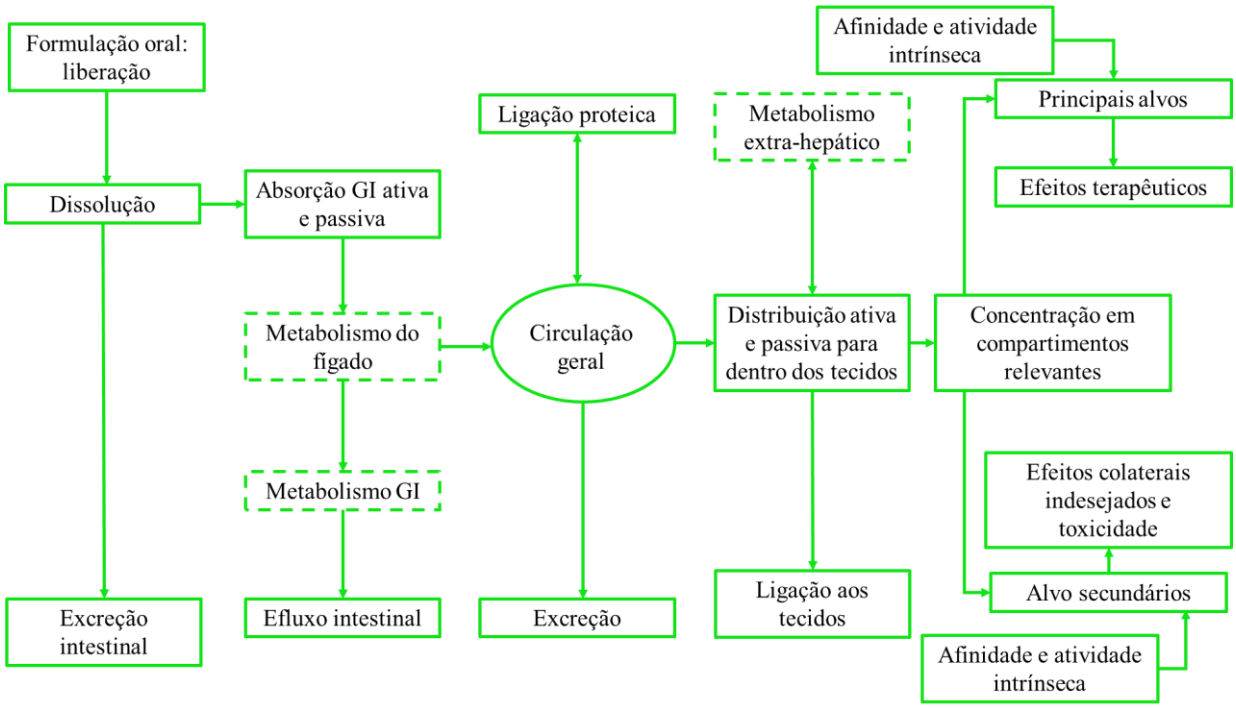
2.2.6.3 Metabolismo

Os citocromos P450 (CYP) são as principais fontes de variabilidade na farmacocinética e na resposta aos medicamentos. De 57 CYPs humanos supostamente funcionais, apenas cerca de uma dúzia de enzimas, pertencentes às famílias CYP1, 2 e 3, são responsáveis pela biotransformação da maioria das substâncias estranhas, incluindo 70-80% de todos os medicamentos em uso clínico [182]. Muitos medicamentos são metabolizados pelas enzimas do citocromo P450 (CYP), predominantemente encontradas no retículo endoplasmático dos hepatócitos, são uma classe de enzimas que catalisam uma gama de biotransformação oxidativa e redutiva, incluindo hidroxilação de carbono, oxidação de heteroátomo, oxidação de ligação, dessaturação de hidrocarboneto e desalogenação de halocarbono, entre outros [183]. A biotransformação de drogas catalisadas por CYP para metabólitos pode resultar em consequências favoráveis ou indesejadas [184]. Favoráveis porque contribuem para diminuir a toxicidade ou neutralizar determinado composto potencialmente perigoso ao organismo. Indesejáveis porque pode diminuir a potencialidade de uma determinado fármaco. Além disso, a inibição dessas isoenzimas é considerada uma das principais causas de interações medicamentosas relacionadas à farmacocinética levando a efeitos adversos ou efeitos tóxicos devido à menor depuração e acúmulo do medicamento ou de seus metabólitos [185].

A P-glicoproteína (P-gp), a glicoproteína de permeabilidade, é um membro da superfamília (ABC) de ligação ao ATP. O P-gp pode ser visto como uma rede de barreira defensiva única contra a entrada de xenobióticos no corpo, o que pode diminuir a biodisponibilidade dos medicamentos administrados evitando seu acúmulo intracelular [186]. Isso indica que a eficácia dos medicamentos é reduzida. Além disso, altera a farmacocinética e farmacodinâmica de seus substratos, alterando as características ADMET. A P-gp é uma proteína de transporte envolvida não apenas na desintoxicação de células, exportando um grande número de toxinas quimicamente não relacionadas, mas também apresenta o fenômeno de multirresistência ocorrendo, entre outros, com agentes anticâncer [187]. Atualmente existe vários modelos *in silico*, *in vitro* e *in vivo* que podem avaliam a capacidade de moléculas serem substratos para a P-gp.

A figura 12 mostra os diversos processos metabólicos envolvidos após a administração oral de uma droga.

Figura 12. Apresentação esquemática do destino de uma droga no corpo após a administração oral. Os processos metabólicos estão em caixas tracejadas e os processos farmacocinéticos estão em caixas com linhas.



Adaptado de Testa, 2014 [176].

2.2.6.4 Toxicidade

A toxicologia *in silico* desempenha um papel vital na avaliação da segurança/toxicidade de produtos químicos e do processo de desenvolvimento de medicamentos. Abordagens computacionais de toxicologia preditiva são usadas em combinação com dados experimentais obtidos a nível molecular, celular, tecidual e em diferentes níveis, possibilitando melhorar a segurança e avaliar as interações entre perigos potenciais com os componentes do sistema biológico [188].

A genetoxicidade ou mutagenicidade é um dos parâmetros mais estudados em toxicologia preditiva e é definida como a capacidade de um agente danificar o DNA ou alterar a sequência de DNA de modo a causar mutação. Os efeitos mais graves dessas mutações são neoplasias, neoplasias hereditárias ou defeitos após o nascimento [189]. No sistema reprodutivo, as consequências são potenciais efeitos adversos na fertilidade, desenvolvimento e comportamento fetal.

2.3 A Bioinformática e o planejamento racional de fármacos

Segundo Luscombe *et al* (2001) [190], a Bioinformática é definida como a aplicação de ferramentas computacionais para organizar, analisar, compreender, visualizar e armazenar informação associada a macromoléculas biológicas. A Bioinformática é uma ciência interdisciplinar que abrange genômica, transcriptômica, proteômica, genética de populações e filogenética molecular [191].

No planejamento racional de fármacos, a Bioinformática pode ser usada para 1) identificar alvos de drogas com atividades importantes através de análise comparativa de sequências e estruturas proteicas; 2) predizer ou refinar alvos que possam atuar sobre o fármaco para alcançar o resultado terapêutico projetado e minimizar os efeitos colaterais e 3) avaliar o potencial de resistência aos medicamentos.

2.3.1 Base de dados de estruturas proteicas 3D

As proteínas são macromoláculas formadas por unidades menores denominadas de aminoácidos que desempenham a maior parte das funções biológicas e por isso, muitos pesquisadores desenvolveram grande interesse neles, pois consistem em novos domínios de conhecimento e aplicação [192]. Experimentos bioquímicos e analíticos produziram uma enorme quantidade de informações relacionadas às proteínas, sobre as sequências de aminoácidos, estruturas tridimensionais de proteínas e, sua função [193]. Com o desenvolvimento de ferramentas computacionais, tornou-se possível compartilhar e vincular essas informações. Assim os bancos de dados constituem uma grande fonte de informações importantes para a investigação de causas de doenças, no desenvolvimento de medicamentos seletivos e compreensão de vias de mecanismo de ação.

A riqueza dos dados proteômicos permite aos pesquisadores fazer perguntas biológicas complexas e obter novas informações científicas. Sites e ferramentas de software de análise de dados foram desenvolvidos para organizar e fornecer anotações biológicas para proteínas com a finalidade de apoiar análises de sequência, estruturais, funcionais e evolutivas no contexto da biologia de vias, redes e sistemas [194].

Os bancos de dados de Bioinformática de proteínas podem ser classificados principalmente como bancos de dados de sequência, dados 2D, estrutura 3D, enzimas e vias, família e de domínio, expressão gênica, genoma, específicos de organismos, bancos de dados filogenômicos, polimorfismo e mutação, interação proteína-proteína, proteômicos, ontologias, entre outros [194]. A tabela 9 mostra os bancos de dados disponíveis de proteínas 3D.

O Banco de Dados de Proteínas (do inglês "*Protein Data Bank"* - PDB) é o maior banco mundial de dados estruturais de macromoléculas biológicas. Foi estabelecido no Brookhaven National Laboratories (BNL) em 1971 como um banco de estruturas cristalinas de macromoleculares biológicas. O modo de acesso aos dados do PDB mudou ao longo dos anos como resultado da tecnologia melhorada, o uso inicial do PDB foi limitado a um pequeno grupo de especialistas envolvidos na pesquisa estrutural. Hoje, são depositadas proteínas obtida de variadas técnicas, dentre elas: determinação da estrutura cristalina de raios-X, RMN, microscopia crioeletrônica e modelagem teórica [195]. Atualmente conta com mais de 177.009 estruturas depositadas.

Tabela 9. Bancos de dados de estruturas proteicas.

Nome abreviado do banco de dados	Nome do banco de dados	Site	Ref.
DisProt	Banco de dados de distúrbios proteicos	http://www.disprot.org/	[196]
MobiDB	Banco de dados de proteínas intrinsecamente desordenadas e móveis	http://mobidb.bio.unipd.it/	[197]
ModBase	Banco de dados de modelos comparativos de estrutura de proteínas	http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi	[198]
PDBe	Banco de dados de proteínas na Europa	http://www.ebi.ac.uk/pdbe/	[199]
PDBj	Protein Data Bank no Japão	http://pdbj.org/	[200]
PDBsum	Banco de dados estruturas proteicas 3D	http://www.ebi.ac.uk/pdbsum/	[201]
ProteinModelPortal	Portal de modelos de proteínas da base de conhecimento de biologia estrutural PSI-Nature	http://www.proteinmodelportal.org/	[202]
RCSB-PDB	Banco de dados de proteínas no	http://www.pdb.org/	[203]

RCSB

Fonte:Sers/tWu *et al*, 2017 [194].

2.3.2 Alinhamento de sequências proteicas

O alinhamento de sequências, como o próprio nome diz, significa alinhar sequências como ferramenta de busca de similaridade entre sequências biológicas (DNA e proteínas) [204]. Esse método desempenha um papel fundamental em várias pesquisas, tais como construção de banco de dados e consulta, previsão da estrutura e função da proteína. Além disso, o alinhamento de sequências proteicas tem sido crucial no campo da Bioinformática para o conhecimento de regiões homólogas compartilhadas por diversas espécies e detecção de resíduos importantes ainda não investigados experimentalmente [205].

No alinhamento, é importante distinguir alguns termos importantes: a similaridade, que é o grau de semelhança entre as propriedades físico-químicas de aminoácidos; a identidade, que se refere aos aminoácidos idênticos presentes nas sequências alinhadas e a homologia, que denota ancestralidade comum. Quanto maior a similaridade e a identidade entre duas ou mais sequências biológicas, maior a semelhança estrutural e funcional entre elas. A Figura 13 mostra o resultado parcial de um alinhamento. À esquerda, o nome das espécies com suas respectivas sequências de aminoácidos alinhadas e à direita, a quantidade de aminoácidos presentes nas sequências. As regiões em cinza correspondem a aminoácidos não idênticos e não similares, as regiões em amarelo correspondem a aminoácidos apenas similares e as regiões em vermelho correspondem a aminoácidos apenas idênticos, os tracejados correspondem a inserções ou deleções de aminoácidos.

Figura 13. Resultado parcial de um alinhamento usando o servidor web Clustal Omega. Cada letra corresponde a um aminoácido da sequência proteica; as regiões em cinza correspondem às regiões não similares e não idênticas entre as sequências alinhadas, as regiões em amarelo correspondem aos aminoácidos similares e as regiões em vermelho às regiões idênticas.

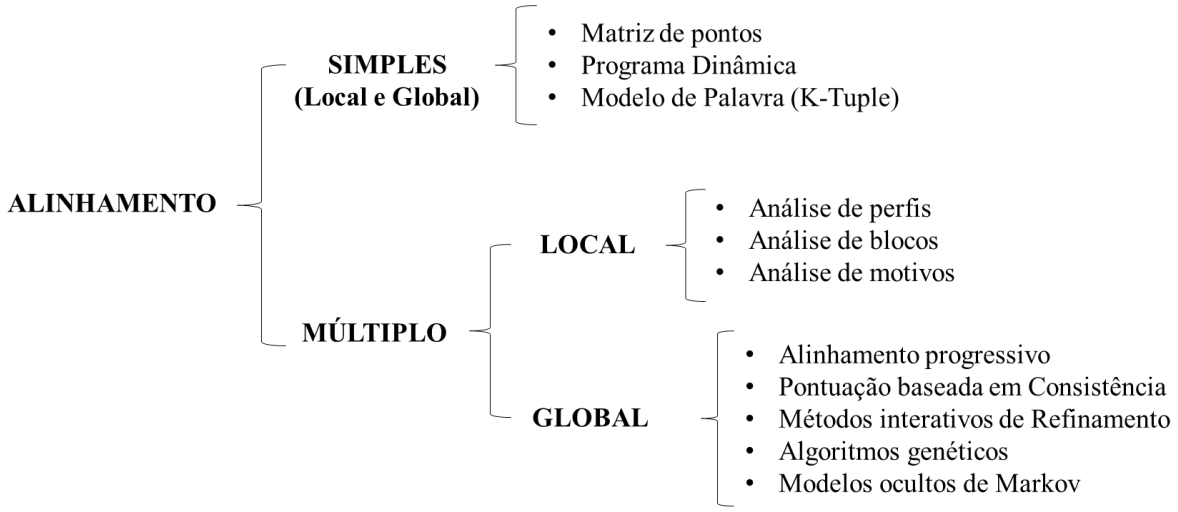
```
L. major          TCINSIGNGLVIDAETES----VVIKPKQGFGGLGGRYVLPTALANINAFYRRCPGK-LI 246
L. donovani       TCINSIGNGLVIDAETES----VVIKPKQGFGGLGGRYVLPTALANVNAFYRRCPGK-LI 247
L. infantum       TCINSIGNGLVIDAETES----VVIKPKQGFGGLGGRYVLPTALANVNAFYRRCPGK-LI 246
L. mexicana       TCINSIGNGLVIDAETES----VVIKPKQGFGGLGGHYILPTALANVNAFYRRCPEK-LI 246
L. braziliensis   TCINSIGNGLVIDVETES----VVIKPKQGFGGLGGRYVFPTALANVNAFYRRCPGK-LI 280
L. amazonensis    TCINSIGNGLVIDAETES----VVIKPKQGFGGLGGHYILPTALANVNAFYRRCPEK-LI 246
T. cruzi          TCVNSVGNGLVIDAESES----VVIKPKQGFGGLGGKYILPTALANVNAFYRRCPDK-LV 247
H. sapiens        -------DGLIVTNTTVSRPAGLQGALRSETGGLSGKPLRDLSTQTIREMYALTQGRVPI 329
```

Fonte: Maia, 2021.

Segundo Junqueira; Braun; Verli (2014) [206], técnicas de alinhamento são fundamentais na construção de algoritmos que visam comparar a informação de diversas sequências biológicas, sendo possível comparar duas ou centenas de sequências, sejam elas de espécies do mesmo gênero ou de gêneros diferentes. Assim, existem dois tipos de alinhamentos: alinhamento simples e múltiplo. Cada um ainda pode ser dividido em alinhamento local e global (Figura 14). Alinhamentos locais comparam apenas regiões similares e parciais, e alinhamentos globais comparam toda a sequência biológica. Alinhamentos simples comparam apenas duas sequências e alinhamento múltiplo alinha diversas sequências.

O principal algoritmo envolvido no processamento de alinhamentos globais foi desenvolvido por Needleman e Wunsch durante a década de 1970, considerado o primeiro algoritmo a aplicar o método de programação dinâmica para a comparação de sequências biológicas. Em 1981, um novo algoritmo desenvolvido por Smith e Waterman, é considerado o principal algoritmo utilizado por programas para realização de alinhamentos locais. A programação dinâmica é o método mais utilizado para realizar o alinhamento de sequências. O método é baseado na solução de problemas complexos através da resolução dos seus diversos subproblemas. A resolução dos subproblemas e seus resultados são armazenados pelo algoritmo. A vantagem consiste em resolver problemas complexos através de uma série de subproblemas. Assim, como o algoritmo acumula os resultados dos diferentes subproblemas, acelera a resolução do problema complexo [206]. Na tabela 10 estão listadas algumas ferramentas web usadas em alinhamentos de sequências simples e múltiplas.

Figura 14. Tipos de alinhamentos e os algoritmos aplicados.



Fonte: Junqueira; Braun; Verli, 2014 [205].

Tabela 10. Alguns softwares e servidores web de alinhamentos de sequências e seu respectivo tipo de alinhamento e algoritmo.

Softwares	Tipo de alinhamento	Algoritmo ou método	Sites
BLAST	Simples/Local	K-Tuples	https://blast.ncbi.nlm.nih.gov/Blast.cgi
CLUSTALW	Múltiplo/Local ou Global	Alinhamento progressivo	https://www.genome.jp/tools-bin/clustalw
T-COFFEE	Múltiplo/Local ou Global	Alinhamento progressivo mais sensível	http://tcoffee.crg.cat/
MUSCLE	Múltiplo/Local ou Global	Alinhamento progressivo/interativo	https://www.ebi.ac.uk/Tools/msa/muscle/
MULTALIN	Múltiplo/Local ou Global	Programação dinâmica/clustering	http://multalin.toulouse.inra.fr/multalin/
Kalign	Múltiplo/Global	Alinhamento progressivo	https://www.ebi.ac.uk/Tools/msa/kalign/
FASTA	Múltiplo/Local ou Global	Método heurístico	https://www.ebi.ac.uk/Tools/sss/fasta/
MAFFT	Múltiplo/Local ou Global	Alinhamento progressivo/interativo	https://www.ebi.ac.uk/Tools/msa/mafft/

2.3.3 Construção tridimensional de proteínas

A determinação experimental é o melhor processo para se obter a estrutura tridimensional de uma proteína, que pode ser obtida por técnicas de Difração de raios-X e RMN. Porém, essas técnicas podem ser custosas e levar muito tempo para se obter a estrutura final. Portanto, o desenvolvimento de métodos computacionais é uma alternativa mais barata, enquanto em alguns casos, a única possibilidade de obtenção de modelos estruturais para algumas proteínas [206].

Segundo pesquisadores do Centro de Previsão de Estrutura de Proteínas dos Estados Unidos, que se reúnem bianualmente em um encontro de caráter mundial denominado CASP (do inglês "*Critical Assessment of Protein Structure Prediction*") (http://predictioncenter.org/index.cgi) para avaliação e competição de predição de estruturas proteicas, uma classificação mais ampla divide os métodos de predição de estrutura de proteínas da seguinte forma (Figura 15):

- Métodos independentes de estruturas moldes. Exemplo: Método *ab initio*;
- Métodos dependentes de estruturas moldes. Exemplo: Método Threading e Modelagem Comparativa ou Modelagem por Homologia.

2.3.3.1 Métodos

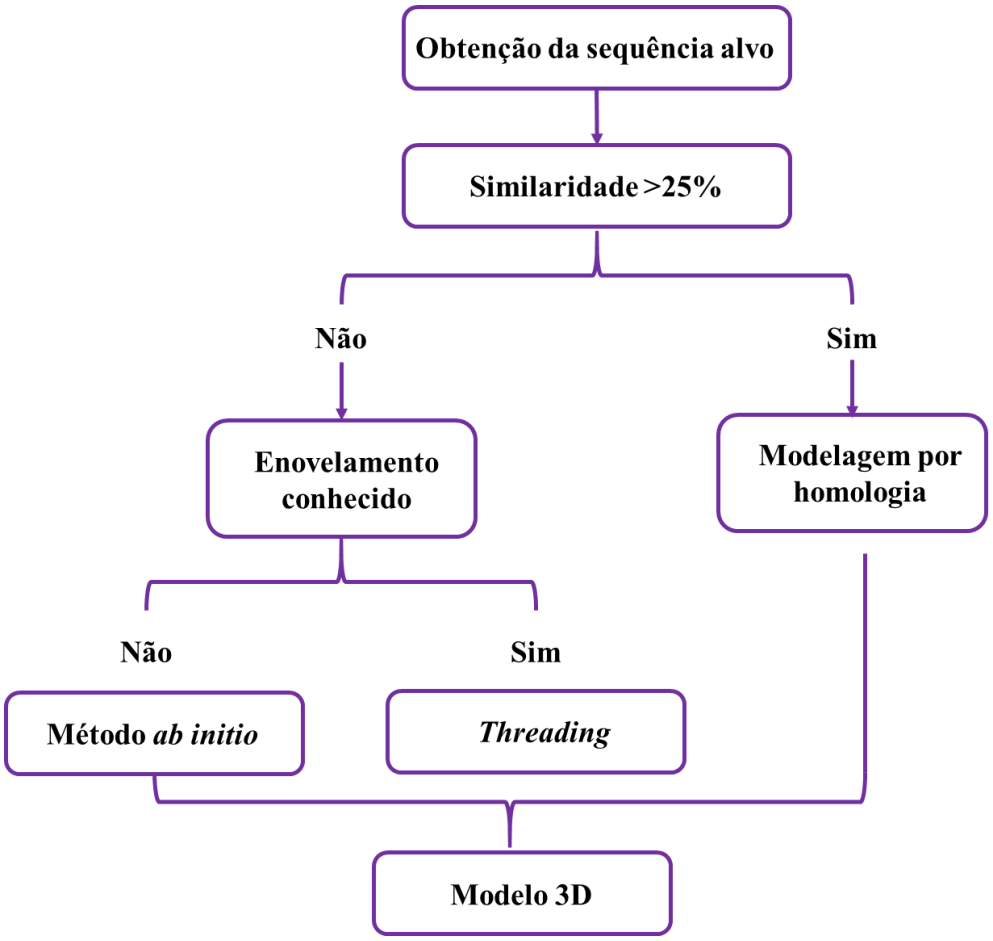
Método ab initio

O método *ab initio* é o único método que não necessita de estruturas moldes para a predição de estruturas proteicas e são fundadas em princípios de termodinâmica e com base na informação de que a estrutura nativa de uma proteína corresponde ao estado termodinamicamente mais estável, ou seja, ao mínimo de energia livre global [207].

Os modelos gerados pelo método *ab initio* são baseados em funções de energia potencial que descrevem a física de um estado conformacional atual e são usadas para pesquisar o estado nativo da estrutura proteica. Informações estruturais são usadas apenas na parametrização de potenciais empíricos de todos os átomos usados no campo de força (funções de potenciais de energia) [207], [208].

Atualmente, a maioria dos métodos de modelagem bem sucedidos, incluindo os métodos livres de moldes (método *ab initio*), de alguma forma usam as informações estruturais do banco de dados do PDB, pelo qual conformações proteicas podem ser construídas rapidamente [209]. A combinação do método *ab initio* usando fragmentos de estruturas proteicas como informação na modelagem de proteínas reduzem a carga computacional significatimente e a confiabilidade da previsão pode ser bastante aprimorada.

Figura 15. Fluxograma para a predição da estrutura tridimensional de uma proteína.



Método Threading

O método de predição do enovelamento ou *threading* parte da ideia de observações de que a estrutura 3D é mais conservada que a sequência, de forma que mesmo sequências com pouca similaridade podem possuir estruturas muito semelhantes, o que limita o número de enovelamentos que proteínas podem assumir. Nesse método, também são usadas proteínas com estruturas 3D conhecidas e depositadas no PDB, de onde as informações sobre os tipos de enovelamento são extraídas e armazenadas em bancos de dados de tipos de enovelamentos. Por ser um método menos dependente da proximidade entre a sequência de aminoácidos da proteína de interesse, pode ser aplicável quando o alinhamento entre a estrutura primária da proteína de interesse e de uma ou mais proteínas de referência (moldes) apresentam uma identidade entre 20% e 30% [206].

Segundo os criadores do método [210], o objetivo é tentar prever a estrutura nativa aproximada da sequência da proteína alvo, alinhando a sequência às estruturas moldes. O alinhamento entre a sequência alvo e a estrutura do modelo é conduzido de forma a maximizar a função de energia/pontuação do alinhamento.

Modelagem por homologia

A modelagem por homologia é o método mais utilizado para predição de estrutura proteica e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde [206]. Consiste no desenho de proteínas a partir de outras proteínas similares e homólogas que já se conhece a estrutura tridimensional por métodos experimentais. Assim, essa metodologia contribui para o conhecimento de proteínas ainda desconhecidas experimentalmente e que podem se tornar importantes alvos terapêuticos [211].

A modelagem por homologia necessita da estrutura de uma proteína conhecida (molde) para gerar uma proteína desconhecida experimentalmente por meio de sua sequência alvo, desde que compartilhe aproximadamente 30% ou mais de semelhança na sequência ou estrutura do modelo [212].

A modelagem comparativa envolve as seguintes etapas: (1) identificação da sequência alvo, (2) alinhamento de sequências alvo e proteínas moldes, e obtenção das estruturas moldes, (3) construção das coordenadas atômicas da estrutura alvo, e (4) avaliação do modelo (Figura 15) [213].

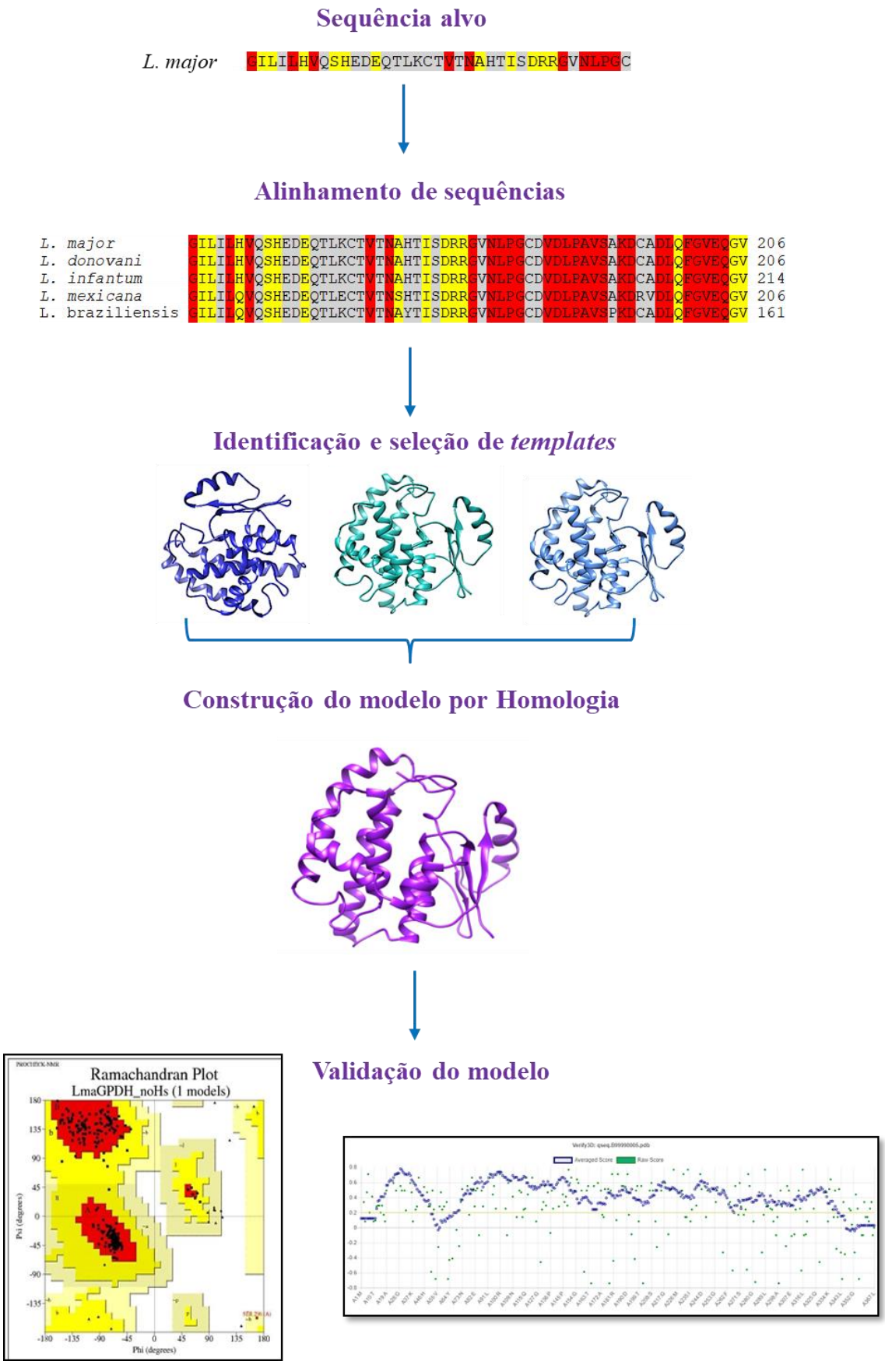
Existem diversos programas de modelagem por homologia representados na tabela 11.

Tabela 11. Ferramentas e servidores web populares para modelagem por homologia [213].

Ferramentas ou servidores	Site	Descrição	Ref.
Modeller	https://www.salilab.org/modeller/	É uma ferramenta gratuita que gera 5 modelos de estruturas 3D de proteínas com o método de restrições espaciais.	[171]
I-Tasser	https://zhanggroup.org//I-TASSER/	É um servidor que fornece um serviço baseado na internet para previsão de proteínas e foi considerado um dos	[214]

		melhores métodos pela CASP.	
Swiss-model	https://swissmodel.expasy.org/	É um servidor que fornece a estrutura 3D da proteína a partir da sequência de seus aminoácidos Ele fornece uma interface da web amigável.	[215]
Molecular Operating Environment (MOE)	https://www.chemcomp.com/Products.htm	É uma combinação de correspondência de segmento e modelagem de inserção ou abordagens de regiões de exclusão.	[216]
Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	Esta modelagem usa várias ferramentas de detecção para gerar estruturas 3D. Ela contém recursos especiais, como previsão de ligação de ligante e análise de variante entre a sequência de aminoácidos da proteína.	[217]
Robetta	http://www.robetta.org/	Com base no método de inserção de fragmento ROSETTA, fornece modelos *ab initio* e homólogos de regiões de proteína.	[218]
Portal do modelo de proteína (PMP)	https://www.proteinmodelportal.org/	PMP fornece interface interativa para construção de modelos e qualidade avaliação.	[219]
ICM	https://www.molsoft.com/homology.html	Seus recursos incluem construção rápida de modelo, loop, previsão, validação de modelo e refinamento.	[220]
Prime	https://www.schrodinger.com/products/prime	É um pacote poderoso para a previsão da estrutura da proteína. Constrói estruturas com alta precisão, fornece simulação avançada. Faz modelagem de homologia e dobra de reconhecimento mesclado em um pacote.	[221]
scwrl4	http://dunbrack.fccc.edu/scwrl4/index.php	É baseado em um novo algoritmo e uma nova função potencial, estimativas de densidade de kernel; média de amostras de conformações; ligação de hidrogênio anisotrópica rápida; algoritmo de decomposição em árvore para resolver o problema combinatório; entre outros.	[222]

Figura 15. Representação das etapas da modelagem por homologia.



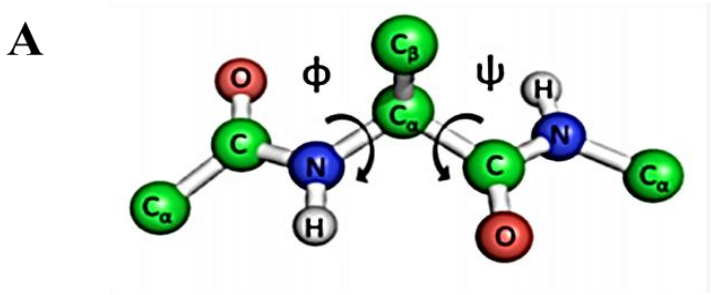
2.3.3.2 Validação dos modelos

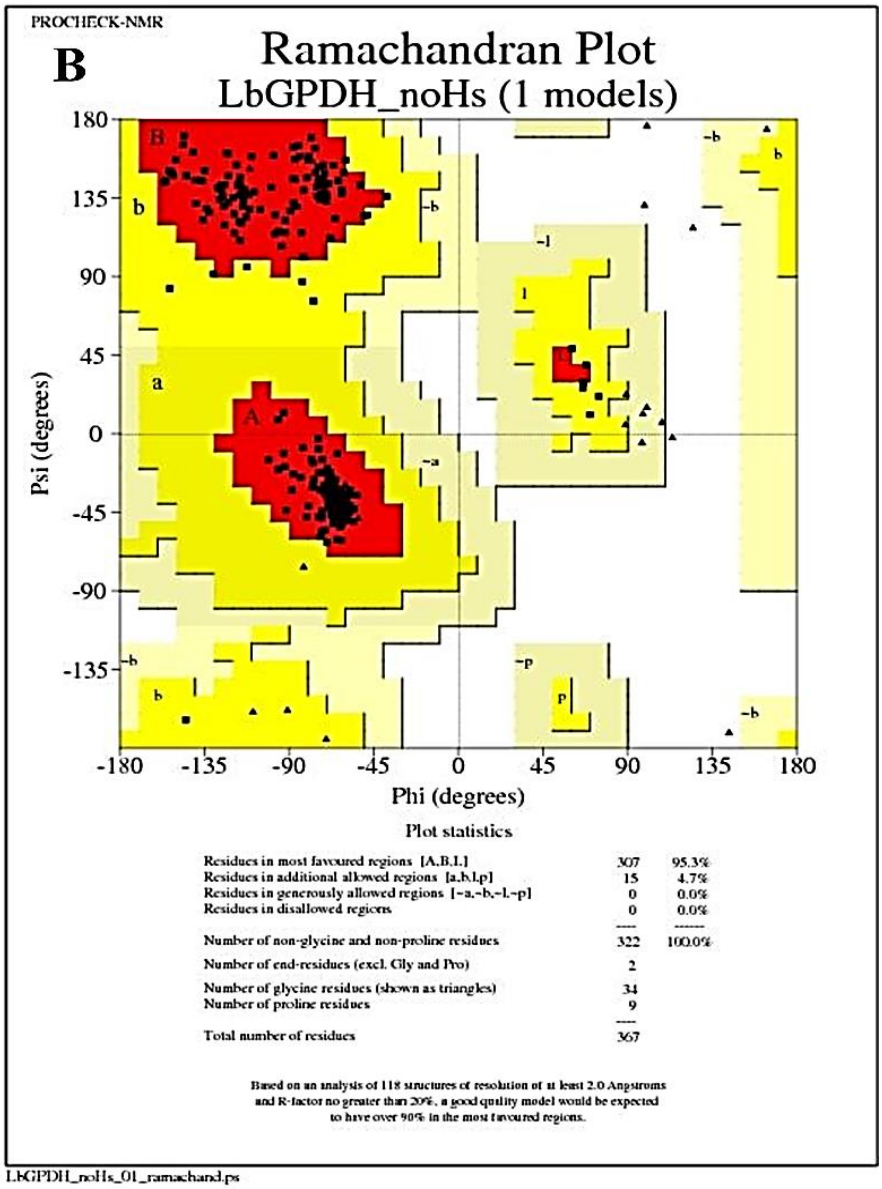
Gráfico de Ramachandran

O gráfico de Ramachandran é um dos principais métodos de avaliação da qualidade estereoquímica de modelos 3D de proteínas. Além disso, é uma ferramenta padrão usada na determinação e definição da estrutura secundária de proteínas obtidas experimentalmente [223]. O Gráfico de Ramachandran consiste na representação de todas as combinações possíveis de ângulos diédricos Ψ (*psi*) versus os φ (*phi*) da cadeia principal dos aminoácidos de um polipeptídeo (Figura 16 A). Para cada resíduo em uma proteína, dois ângulos de torção φ e ψ determinam a conformação do esqueleto da proteína, exceto para os aminoácidos prolina e glicina que podem formar mais de dois tipos de torções [224].

As "regiões favorecidas" destacadas em vermelho na figura 16 B incluem as conformações que dão origem a α hélices e folhas β, respectivamente. As "regiões permitidas", destacadas em amarelo e bege na figura 16 B correspondem às conformações de impedimento estérico ou loops. Essas regiões são "permitidas" no sentido de que quando os átomos do peptídeo recebem raios padrão, eles não colidem. Uma região adicional, às vezes chamada de região da ponte (porque faz a ponte entre as regiões alfa e beta) torna-se permitida se o átomos recebem raios menores que representam os menores valores que poderiam ser considerados plausíveis [225]. Enquanto as "regiões proibidas" que correspondem a parte branca da figura 16 B, estão relacionados a tipos de torções inexistentes. Portanto, o aminoácido que estiver na parte branca do gráfico de Ramachandran, signica um resíduo que não foi corretamente modelado e está incorreto, exceto para os aminoácidos prolina e glicina. Cada ponto preto no gráfico de Ramachandran (Figura 16 B) representa um aminoácido. Pontos vermelhos na parte branca do gráfico significam resíduos incorretos. Assim, para que uma estrutura tridimensional de proteína seja validada através do gráfico de Ramachandran, é necessário que as regiões favorecidas e permitidas totalizem 90% de resíduos presentes nessas regiões [226].

Figura 16. Gráfico de Ramachandran. A) Modelo de bola e bastão de um dipeptídeo com o resíduo Ala acentral indicando as rotações definidas pelos ângulos de torção de φ e ψ. B) O gráfico Ramachandran com contornos definindo as regiões classicamente favorecidas (destacado em vermelho), permitidas (destacado em amarelo e bege) e proibidas (região branca). Os pontos pretos indicam cada aminoácido presente no modelo da proteína GPDH de *Leishmania braziliensis* obtida através de modelagem por homologia.



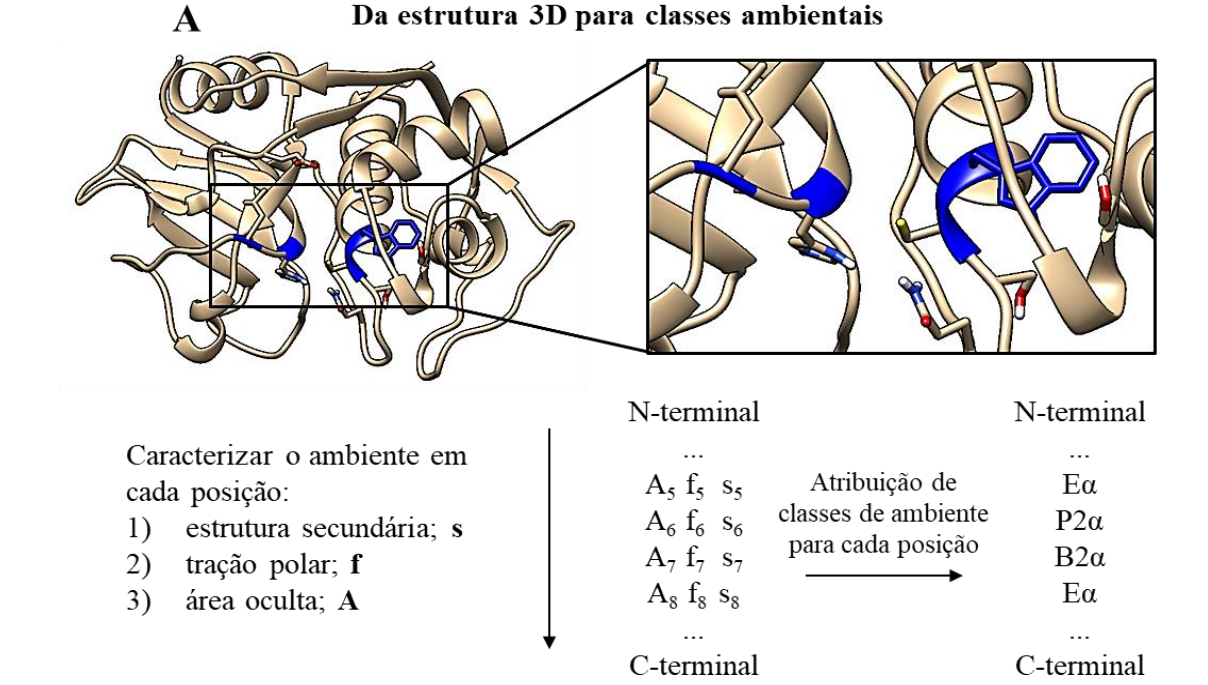


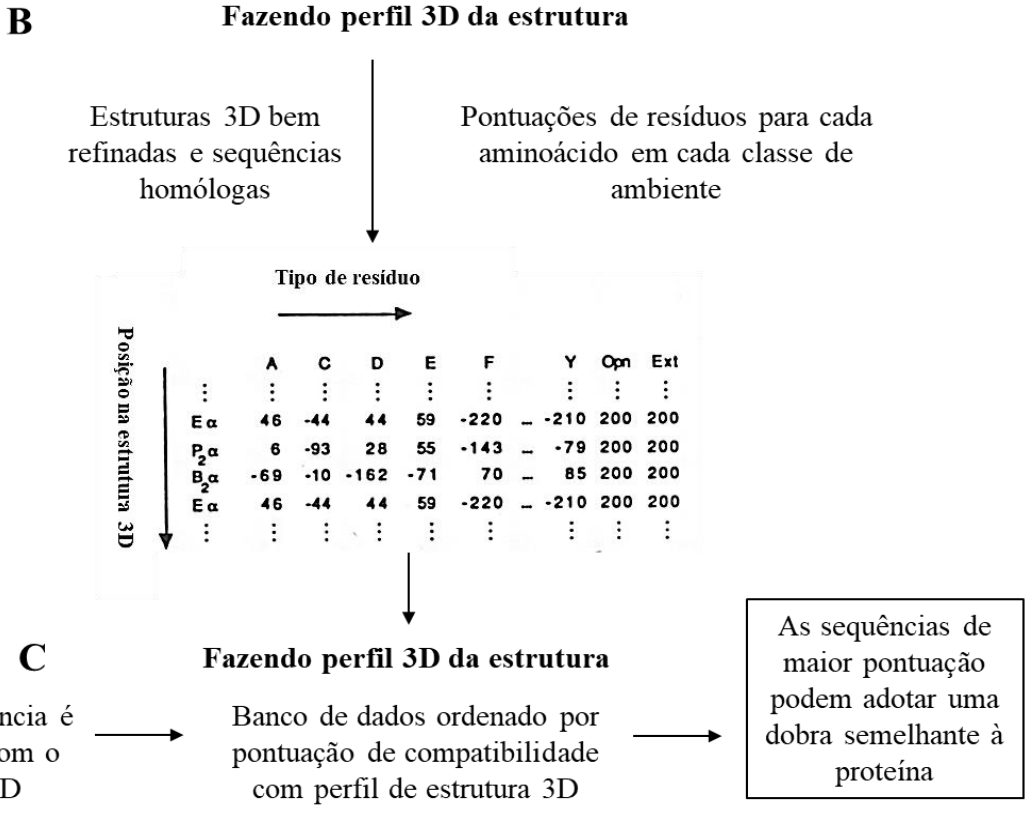
Verify 3D

O Verify 3D é uma ferramenta web (https://servicesn.mbi.ucla.edu/Verify3D/) que determina a compatibilidade do perfil tridimensional (3D) de uma proteína com sua própria sequência de aminoácidos (1D) atribuindo uma classe estrutural com base em sua localização e ambiente (alfa, beta, loop, polar, apolar, etc.) e comparando os resultados com boas estruturas. Assim, a exatidão de um modelo de proteína pode ser verificada por seu perfil 3D, independentemente de o modelo ter sido derivado por raios-X, ressonância magnética nuclear (NMR) ou procedimentos computacionais [227].

Segundo Bowie; Ltcy; Eisenberg (1991) [228], o método, descrito na figura 17, começa com uma estrutura 3D conhecida e determina três características do ambiente de cada resíduo: (i) a área total da cadeia lateral que está oculta por outros átomos da proteína; (ii) a fração da área da cadeia lateral que é coberta por átomos polares ou água; e (iii) a estrutura secundária local. Com base nestes parâmetros, cada posição de resíduo é categorizada em um ambiente. Com isso, uma estrutura de proteína 3D é convertida em 1D, como uma sequência, que representa a classe de ambiente de cada resíduo na estrutura da proteína dobrada.

Figura 17. Descrição esquemática da construção de um perfil de estrutura 3D (A e B) e de uma busca de compatibilidade 3D do banco de dados de sequência (C). O perfil da estrutura 3D da Cruzaína de *Trypanossoma cruzi* na parte inferior de (A) é uma parte do perfil para a, dando pontuações para apenas quatro posições da estrutura (correspondendo aos resíduos 5, 6, 7 e 8). No quadro de pontuação (B), Opn representa as penalidades gap relativamente baixa e Ext, penalidades de extensão de lacuna.





Bowie; Ltcy; Eisenberg (1991) [228].

2.3.4 Resistência à fármacos

O estudo de características genéticas de determinadas proteínas candidatas a alvos abre a possibilidade para estudar a evolução e a ação contra a resistência aos medicamentos. Um dos principais mecanismos de resistência à medicamentos observado é baseado numa afinidade diminuída da enzima alvo para o composto causado por mutações específicas em seus genes [229].

Embora parâmetros clínicos continuem a serem usados para orientar o tratamento de drogas, deve-se também considerar os fatores genéticos, que podem ser responsáveis por até 40% das diferenças interindividuais na resposta e metabolismo da droga [230]. Fatores genéticos têm demonstrado ser a influência mais importante sobre os resultados do tratamento para certas drogas e classes de medicamentos. Variações genéticas podem alterar a estrutura de uma proteína alvo por meio de mutações na região de codificação do gene ou a quantidade de proteína expressa modulando a regulação de genes. Mudanças estruturais em receptores e enzimas podem afetar a interação droga-enzima e, consequentemente, a resposta à droga [231].

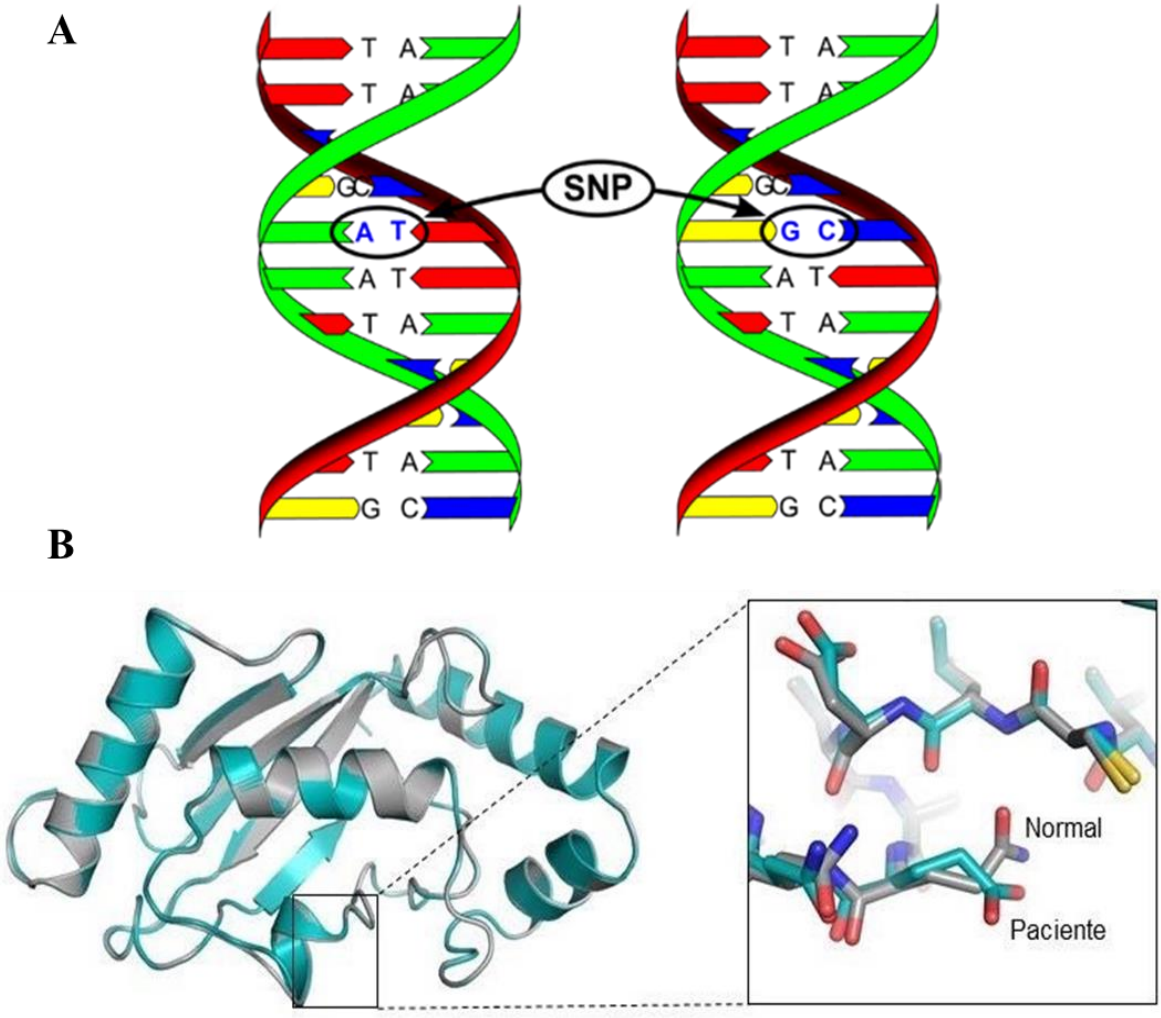
Segundo Marin *et al* (2012) [232], as mudanças nos genes que codificam os transportadores de soluto podem ser responsáveis por uma considerável variabilidade interindividual na absorção do fármaco e pela falta de sensibilidade aos substratos desses transportadores. Além disso, alterações nas proteínas envolvidas na exportação de drogas podem afetar sua localização subcelular e capacidade de transporte. Em relação à ativação ou inativação de pró-fármacos, as variações genéticas afetam a depuração do fármaco e podem determinar a falta de resposta terapêutica. Também pode diminuir a sensibilidade a agentes farmacológicos que atuam por meio de alvos moleculares ou vias de sinalização.

2.3.4.1 Polimorfismo de nucleotídeo único

Polimorfismos de nucleotídeo único (do inglês *Single nucleotide polymorphisms* - SNP) são mutações em um único par de bases no DNA genômico com sequências de diferentes alternativas existentes em indivíduos de alguma população, em que o alelo menos frequente tem uma abundância de 1% ou superior [233]. A figura 18 mostra a representação

de um SNP, que consite na troca de um nucleotídeo por outro com uma frequência mínima de 1% na população. No caso da figura 18, num indivíduo, no gene X, um determinado nucleotídeo numa determinada posição ocorre uma adenina (A) e no outro indivíduo, o mesmo gene e na mesma posição, a adenina é substituída por uma guanina (G). Essa mudança em um único nucleotídeo muda a leitura da trinca de nucleotídeos pelos ribossomos levando a alteração do tipo de aminoácido. A mudança de um único aminoácido pode acarretar na produção de uma proteína truncada, proteína não funcional ou aumentar a expressão de uma determinada proteína. Essas consequências podem levar a diversas patologias.

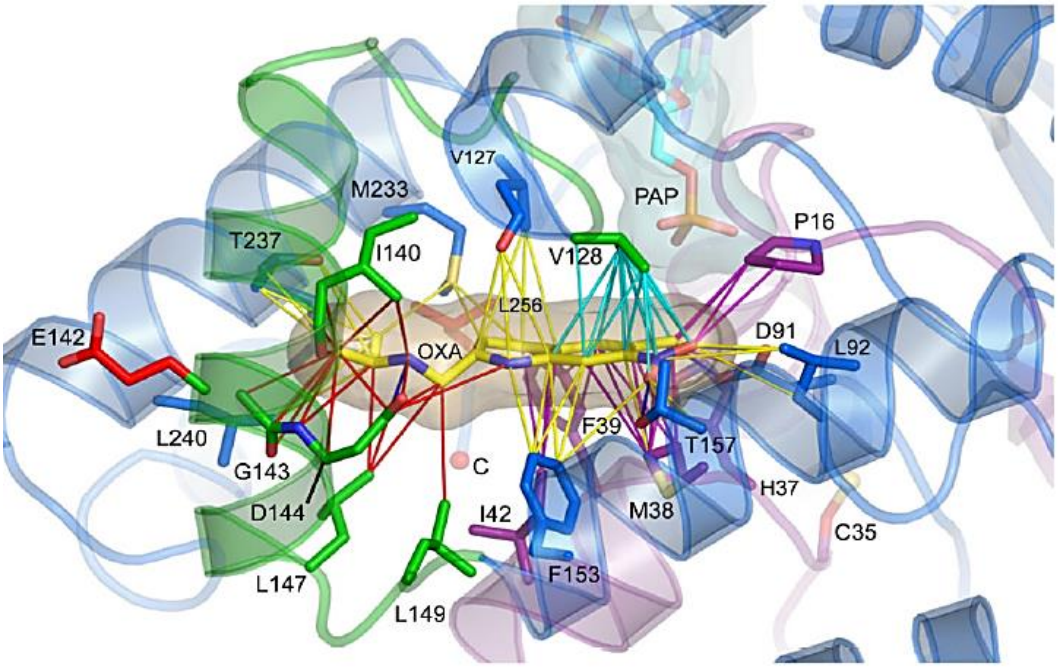
Figura 18. Representação de um polimorfismo de nucleotídeo único e sua consequência. A - SNP e B – troca de aminoácido provocado pelo SNP.



Fonte: https://mydnahealth.co.uk/national-dna-day/ (Acesso: 17 de Julho de 2021).

Diversos estudos têm relatado a associação de SNP à susceptibilidade de diversas doenças e resistências à medicamentos [234], [235], [236], [237], [238]. Um exemplo de SNP que leva à resistência de medicamento é relatado por Valentin *et al* (2013) [239]. Segundo os autores, a Oxamniquine (OXA) foi o fármaco de primeira linha no Brasil até o final dos anos 90 contra a esquistossomose e permaneceu até 2010. A resistência a esse medicamento tem uma base recessiva e resulta em uma redução de aproximadamente 500 vezes na sensibilidade à droga. As bases moleculares para a resistência a OXA são perturbações induzidas pela mutação na estrutura da enzima que impede a ligação e/ou sulfonação da OXA. A substituição da cisteína por arginina na posição 35 (C35R) da enzima sulfotransferase leva a interrupção de aproxidamente 26 interações (Figura 19).

Figura 19. Contatos provavelmente interrompidos pela mutação C35R representados pelas linhas roxas. Os contatos ciano podem estar comprometidos devido a troca de aminoácido.



Fonte: Valentin et al (2013) [239].

3. OBJETIVOS

3.1 Objetivo geral

Avaliar o potencial farmacológico de lignanas e neolignanas frente a doenças negligenciadas (leishmaniose, doença de Chagas e esquistossomose) e neurodegenerativas (Alzheimer) utilizando recursos e abordagens computacionais da Quimio e Bioinformática, bem como a validação experimental.

3.2 Objetivos específicos

- Selecionar lignanas e neolignanas de bancos de dados para serem usados nas abordagens computacionais e investigativas sobre o potencial farmacológico relacionadas às doenças propostas;
- Realizar triagem de lignanas e neolignanas através de propriedades ADMET;
- Utilizar ferramentas de bioinformática para alinhar e avaliar sequências das enzimas alvos para construção de estruturas proteicas tridimensionais;
- Identificar aminoácidos conservados e o sítio ativo de modelos de proteínas teóricos entre as espécies de interesse através de alinhamento de sequências proteicas;
- Construir modelos moleculares teóricos através de modelagem por homologia das proteínas selecionadas, uma vez que estruturas tridimensionais de algumas enzimas ainda não foram determinadas para algumas espécies de *Leishmania*, e possibilitar os estudos de docking molecular;
- Calcular descritores moleculares dos conjuntos de estruturas químicas selecionados;
- Construir modelos preditivos de QSAR contra importantes alvos e espécies de parasitas; e alvos envolvidos direta e indiretamente na via do estresse oxidativo em humanos;
- Realizar a análise preditiva de atividade biológica dos compostos investigados;
- Realizar com auxílio de diferentes softwares e funções de pontuação, o docking molecular, de maneira a permitir selecionar o confôrmero de menor energia que então será submetido às demais metodologias;
- Utilizar abordagens consenso de docking molecular para aumentar a probabilidade de resultados verdadeiros positivos;

- Identificar o modelo de ligação mais provável e analisar as interações dos ligantes com as enzimas;
- Realizar abordagens consenso de triagem virtual baseada no ligante e na estrutura;
- Identificar lignanas e neolignanas *multi-targets*;
- Investigar a flexibilidade de interações dos melhores resultados de complexos proteína-ligante através de análises de simulações de dinâmica molecular;
- Analisar a afinidade de ligação na presença de solvente através de cálculos de energia livre, a partir dos melhores resultados de complexos proteína-ligante obtidos pelas triagens;
- Avaliar a possibilidade de resistência aos compostos selecionados através de análise de polimorfismos de nucleotídeo único dos genes das proteínas alvos;
- Isolar lignanas e neolignanas através de parceria com o Laboratório Multiusuário de Análise e Caracterização, da Universidade Federal da Paraíba;
- Testar lignanas e neolignanas contra formas celulares de *Leishmania major*, *Lesihmania braziliensis* e *Trypanossoma cruzi* em parceria com o Laboratório de Doenças Infecsiosas, da Universidade Federal Delta do Parnaíba, Piauí.

REFERÊNCIAS

[1] J. B. Calixto, “The role of natural products in modern drug discovery.,” *An. Acad. Bras. Cienc.*, vol. 91, pp. 1–7, 2019, doi: 10.1590/0001-3765201920190105.

[2] K. Santana *et al.*, “Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products,” *Front. Chem.*, vol. 9, no. April, 2021, doi: 10.3389/fchem.2021.662688.

[3] A. G. Atanasov *et al.*, “Natural products in drug discovery: advances and opportunities,” *Nat. Rev. Drug Discov.*, vol. 20, no. 3, pp. 200–216, 2021, doi: 10.1038/s41573-020-00114-z.

[4] D. J. Newman and G. M. Cragg, “Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019,” *J. Nat. Prod.*, vol. 83, no. 3, pp. 770–803, 2020, doi: 10.1021/acs.jnatprod.9b01285.

[5] M. A. G. da Silva, “The Importance of Otto Gottlieb for Natural Product Chemistry and Brazilian Biodiversity,” *J. Clin. Biomed. Res.*, vol. 7, no. 366, pp. 1–2, 2019, doi: 10.47363/jcbr/2019(1)107.

[6] C. Caparica, “Otto Gottlieb impulsionou a química de produtos naturais,” *Cienc. Cult.*, vol. 64, no. 1, pp. 10–11, 2012, doi: 10.21800/s0009-67252012000100005.

[7] W. R., M. L. A. e Silva, R. C. Sola Veneziani, S. Ricardo, and J. Kenupp, “Lignans: Chemical and Biological Properties,” *Phytochem. - A Glob. Perspect. Their Role Nutr. Heal.*, no. March, 2012, doi: 10.5772/28471.

[8] F. Zálešák, D. J. Y. D. Bon, and J. Pospíšil, “Lignans and Neolignans: Plant secondary metabolites as a reservoir of biologically active substances,” *Pharmacol. Res.*, vol. 146, no. May, p. 104284, 2019, doi: 10.1016/j.phrs.2019.104284.

[9] M. Gordaliza, P. A. García, J. M. Miguel Del Corral, M. A. Castro, and M. A. Gómez-Zurita, “Podophyllotoxin: Distribution, sources, applications and new cytotoxic derivatives,” *Toxicon*, vol. 44, no. 4, pp. 441–459, 2004, doi: 10.1016/j.toxicon.2004.05.008.

[10] WHO, “First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases,” *World Heal. Organ.*, pp. 1–184, 2010, doi: 10.1177/1757913912449575.

[11] K. Hofstraat and W. H. Van Brakel, “Social stigma towards neglected tropical diseases: A systematic review,” *Int. Health*, vol. 8, no. Suppl 1, pp. i53–i70, 2015, doi: 10.1093/inthealth/ihv071.

[12] E. Von Stebut, “Leishmaniasis,” *JDDG - J. Ger. Soc. Dermatology*, vol. 13, no. 3, pp. 191–200, 2015, doi: 10.1111/ddg.12595.

[13] S. Srivastava, P. Shankar, J. Mishra, and S. Singh, “Possibilities and challenges for developing a successful vaccine for leishmaniasis,” *Parasites and Vectors*, vol. 9, no. 1, pp. 1–15, 2016, doi: 10.1186/s13071-016-1553-y.

[14] D. Savoia, “Recent updates and perspectives on leishmaniasis,” *J. Infect. Dev. Ctries.*, vol. 9, no. 6, pp. 588–596, 2015, doi: 10.3855/jidc.6833.

[15] K. C. F. Lidani *et al.*, “Clinical and epidemiological aspects of chronic chagas disease from southern brazil,” *Rev. Soc. Bras. Med. Trop.*, vol. 53, pp. 1–10, 2020, doi: 10.1590/0037-8682-0559-2019.

[16] L. Gaspar *et al.*, “Current and Future Chemotherapy for Chagas Disease,” *Curr. Med. Chem.*, vol. 22, no. 37, pp. 4293–4312, 2015, doi: 10.2174/0929867322666151015120804.

[17] A. M. Jansen, S. C. D. C. Xavier, and A. L. R. Roque, “Trypanosoma cruzi transmission in the wild and its most important reservoir hosts in Brazil,” *Parasites and Vectors*, vol. 11, no. 1, pp. 1–25, 2018, doi: 10.1186/s13071-018-3067-2.

[18] E. Leger and J. P. Webster, “Hybridizations within the Genus Schistosoma: Implications for evolution, epidemiology and control,” *Parasitology*, vol. 144, no. 1, pp. 65–80, 2017, doi: 10.1017/S0031182016001190.

[19] P. T. Loverde, “Schistosomiasis,” *Adv Exp Med Biol*, 2019, doi: 10.1007/978-3-030-18616-6.

[20] A. C. Zoni, L. Catalá, and S. K. Ault, “Schistosomiasis Prevalence and Intensity of Infection in Latin America and the Caribbean Countries, 1942-2014: A Systematic Review in the Context of a Regional Elimination Goal,” *PLoS Negl. Trop. Dis.*, vol. 10, no. 3, pp. 1–22, 2016, doi: 10.1371/journal.pntd.0004493.

[21] J. H. No, “Visceral leishmaniasis: Revisiting current treatments and approaches for future discoveries,” *Acta Trop.*, vol. 155, pp. 113–123, 2016, doi: 10.1016/j.actatropica.2015.12.016.

[22] C. J. Alvarez VE1, Niemirowicz GT, “Metacaspases, Autophagins and Metallocarboxypeptidases: Potential New Targets for Chemotherapy of the Trypanosomiases.,” *Curr. Med. Chem.*, vol. 20, no. 25, pp. p3069-3077, 2013.

[23] R. Viotti *et al.*, “Side effects of benznidazole as treatment in chronic Chagas disease: Fears and realities,” *Expert Rev. Anti. Infect. Ther.*, vol. 7, no. 2, pp. 157–163, 2009, doi: 10.1586/14787210.7.2.157.

[24] M. Merrifield *et al.*, “Advancing a vaccine to prevent human schistosomiasis,” *Vaccine*, vol. 34, no. 26, pp. 2988–2991, 2016, doi: 10.1016/j.vaccine.2016.03.079.

[25] B. N. Dugger and D. W. Dickson, “Pathology of neurodegenerative diseases,” *Cold Spring Harb. Perspect. Biol.*, vol. 9, no. 7, pp. 1–22, 2017, doi: 10.1101/cshperspect.a028035.

[26] G. G. Kovacs, *Concepts and classification of neurodegenerative diseases*, 1st ed., vol. 145. Elsevier B.V., 2018.

[27] S. P. Gupta, “Advances in Studies on Neurodegenerative Diseases and their Treatments,” *Curr. Top. Med. Chem.*, vol. 20, no. 13, pp. 1141–1141, 2020, doi: 10.2174/1568026620132005190750 27.

[28] J. A. Soria Lopez, H. M. González, and G. C. Léger, “Alzheimer’s disease,” *Handb.*

Clin. Neurol., vol. 167, pp. 231–255, 2019, doi: 10.1016/B978-0-12-804766-8.00013-3.

[29] R. Rajasekaran and Y. P. P. Chen, "Potential therapeutic targets and the role of technology in developing novel antileishmanial drugs," *Drug Discov. Today*, vol. 20, no. 8, pp. 958–968, 2015, doi: 10.1016/j.drudis.2015.04.006.

[30] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: Advances and applications," *Adv. Appl. Bioinforma. Chem.*, vol. 8, no. 1, pp. 37–47, 2015, doi: 10.2147/AABC.S70333.

[31] M. M. Jaghoori, B. Bleijlevens, and S. D. Olabarriaga, "1001 Ways to run AutoDock Vina for virtual screening," *J. Comput. Aided. Mol. Des.*, vol. 30, no. 3, pp. 237–249, 2016, doi: 10.1007/s10822-016-9900-9.

[32] D. Barker, "Lignans," *Molecules*, vol. 24, no. 7, pp. 2–5, 2019, doi: 10.3390/molecules24071424.

[33] V. A. De Souza, C. V. Nakamura, and A. G. Corrêa, "Antichagasic activity of lignans and neolignans," *Rev. Virtual Quim.*, vol. 4, no. 3, pp. 197–207, 2012, doi: 10.5935/1984-6835.20120017.

[34] H. Satake, T. Koyama, S. E. Bahabadi, E. Matsumoto, E. Ono, and J. Murata, "Essences in metabolic engineering of lignan biosynthesis," *Metabolites*, vol. 5, no. 2, pp. 270–290, 2015, doi: 10.3390/metabo5020270.

[35] R. B. Teponno, S. Kusari, and M. Spiteller, *Recent advances in research on lignans and neolignans*, vol. 33, no. 9. 2016.

[36] L. I. Pilkington, "Lignans: A chemometric analysis," *Molecules*, vol. 23, no. 7, pp. 1–24, 2018, doi: 10.3390/molecules23071666.

[37] Q. V. Vo, P. C. Nam, M. Van Bay, N. M. Thong, N. D. Cuong, and A. Mechler, "Density functional theory study of the role of benzylic hydrogen atoms in the antioxidant properties of lignans," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-30860-5.

[38] W. H. Xu, P. Zhao, M. Wang, and Q. Liang, "Naturally occurring furofuran lignans: structural diversity and biological activities," *Nat. Prod. Res.*, vol. 33, no. 9, pp. 1357–1373, 2019, doi: 10.1080/14786419.2018.1474467.

[39] D. D. Dong *et al.*, "Diverse lignans with anti-inflammatory activity from Urceola rosea," *Fitoterapia*, vol. 134, no. January, pp. 96–100, 2019, doi: 10.1016/j.fitote.2019.02.011.

[40] Y. N. Yang, X. Y. Huang, Z. M. Feng, J. S. Jiang, and P. C. Zhang, "Hepatoprotective activity of twelve novel 7′-Hydroxy lignan glucosides from arctii fructus," *J. Agric. Food Chem.*, vol. 62, no. 37, pp. 9095–9102, 2014, doi: 10.1021/jf501859x.

[41] S. Tannous, T. Haykal, J. Dhaini, M. H. Hodroj, and S. Rizk, "The anti-cancer effect of flaxseed lignan derivatives on different acute myeloid leukemia cancer cells," *Biomed. Pharmacother.*, vol. 132, p. 110884, 2020, doi: 10.1016/j.biopha.2020.110884.

[42] P. García-Huertas, F. Olmo, M. Sánchez-Moreno, J. Dominguez, R. Chahboun, and O.

Triana-Chávez, “Activity in vitro and in vivo against Trypanosoma cruzi of a furofuran lignan isolated from Piper jericoense,” *Exp. Parasitol.*, vol. 189, pp. 34–42, 2018, doi: 10.1016/j.exppara.2018.04.009.

[43] S. S. Grecco *et al.*, “Neolignans from leaves of Nectandra leucantha (Lauraceae) display in vitro antitrypanosomal activity via plasma membrane and mitochondrial damages,” *Chem. Biol. Interact.*, vol. 277, pp. 55–61, 2017, doi: 10.1016/j.cbi.2017.08.017.

[44] K. Sowndhararajan, P. Deepa, M. Kim, S. J. Park, and S. Kim, “An overview of neuroprotective and cognitive enhancement properties of lignans from Schisandra chinensis,” *Biomed. Pharmacother.*, vol. 97, no. October 2017, pp. 958–968, 2018, doi: 10.1016/j.biopha.2017.10.145.

[45] M. Maleck, P. de O. Hollanda, M. T. Serdeiro, R. O. de A. Soares, N. A. Honório, and C. G. Silva, “Toxicity and Larvicidal Activity of Podophyllum-Based Lignans Against Aedes aegypti (Diptera: Culicidae),” *J. Med. Entomol.*, vol. 54, no. 1, pp. 159–166, 2017, doi: 10.1093/jme/tjw130.

[46] K. M. Oksman-Caldentey and D. Inzé, “Plant cell factories in the post-genomic era: New ways to produce designer secondary metabolites,” *Trends Plant Sci.*, vol. 9, no. 9, pp. 433–440, 2004, doi: 10.1016/j.tplants.2004.07.006.

[47] M. Yousefzadi *et al.*, “Podophyllotoxin: Current approaches to its biotechnological production and future challenges,” *Eng. Life Sci.*, vol. 10, no. 4, pp. 281–292, 2010, doi: 10.1002/elsc.201000027.

[48] S. Nikolov *et al.*, “Exploitation of the bulgarian flora’s biodiversity as a source of immunomodulatory and/or antineoplastic agents: Current challenges and perspectives,” *Biotechnol. Biotechnol. Equip.*, vol. 21, no. 4, pp. 471–477, 2007, doi: 10.1080/13102818.2007.10817497.

[49] T. Gasteiger, Johann; Engel, *Chemoinformatics: A Textbook*, KGaA: Wein. 2003.

[50] V. M. Alves, R. C. Braga, E. N. Muratov, and C. Horta, “Quimioinformática: Uma introdução.,” *Quim. Nov.*, vol. 41, no. 2, pp. 202–212, 2018.

[51] E. Piccirillo and A. T. Do Amaral, “Virtual screening of bioactive compounds: Concepts and aplications.,” *Quim. Nova*, vol. 41, no. 6, pp. 662–677, 2018, doi: 10.21577/0100-4042.20170210.

[52] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, “Computational methods in drug discovery,” *Pharmacol. Rev.*, vol. 66, no. 1, pp. 334–395, 2014, doi: 10.1124/pr.112.007336.

[53] S. P. Leelananda and S. Lindert, “Computational methods in drug discovery,” *Beilstein J. Org. Chem.*, vol. 12, pp. 2694–2718, 2016, doi: 10.3762/bjoc.12.267.

[54] J. Yang, D. Wang, C. Jia, M. Wang, G. Hao, and G. Yang, “Freely Accessible Chemical Database Resources of Compounds for In Silico Drug Discovery,” *Curr. Med. Chem.*, vol. 26, no. 42, pp. 7581–7597, 2018, doi: 10.2174/0929867325666180508100436.

[55] H. E. Pence and A. Williams, “Chemspider: An online chemical information resource,”

J. Chem. Educ., vol. 87, no. 11, pp. 1123–1124, 2010, doi: 10.1021/ed100697w.

[56] J. Hastings *et al.*, "ChEBI in 2016: Improved services and an expanding collection of metabolites," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1214–D1219, 2016, doi: 10.1093/nar/gkv1031.

[57] S. Kim *et al.*, "PubChem substance and compound databases," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1202–D1213, 2016, doi: 10.1093/nar/gkv951.

[58] A. Gaulton *et al.*, "ChEMBL: A large-scale bioactivity database for drug discovery," *Nucleic Acids Res.*, vol. 40, no. D1, pp. 1100–1107, 2012, doi: 10.1093/nar/gkr777.

[59] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities," *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 198–201, 2007, doi: 10.1093/nar/gkl999.

[60] K. P. Seiler *et al.*, "ChemBank: A small-molecule screening and cheminformatics resource database," *Nucleic Acids Res.*, vol. 36, no. SUPPL. 1, pp. 351–359, 2008, doi: 10.1093/nar/gkm843.

[61] R. Wang, X. Fang, Y. Lu, and S. Wang, "The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures," *J. Med. Chem.*, vol. 47, no. 12, pp. 2977–2980, 2004, doi: 10.1021/jm030580l.

[62] D. S. Wishart, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Res.*, vol. 34, no. 90001, pp. D668–D672, 2006, doi: 10.1093/nar/gkj067.

[63] P. Banerjee, J. Erehman, B. O. Gohlke, T. Wilhelm, R. Preissner, and M. Dunkel, "Super Natural II-a database of natural products," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D935–D939, 2015, doi: 10.1093/nar/gku886.

[64] M. T. Scotti *et al.*, "Sistematx, an online web-based cheminformatics tool for data management of secondary metabolites," *Molecules*, vol. 23, no. 1, pp. 1–10, 2018, doi: 10.3390/molecules23010103.

[65] J. Gu, Y. Gui, L. Chen, G. Yuan, H. Z. Lu, and X. Xu, "Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology," *PLoS One*, vol. 8, no. 4, pp. 1–10, 2013, doi: 10.1371/journal.pone.0062839.

[66] F. Ntie-Kang *et al.*, "CamMedNP: Building the Cameroonian 3D structural natural products database for virtual screening," *BMC Complement. Altern. Med.*, vol. 13, 2013, doi: 10.1186/1472-6882-13-88.

[67] C. Y. C. Chen, "TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening In Silico," *PLoS One*, vol. 6, no. 1, pp. 1–5, 2011, doi: 10.1371/journal.pone.0015939.

[68] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, "ZINC: A free tool to discover chemistry for biology," *J. Chem. Inf. Model.*, vol. 52, no. 7, pp. 1757–1768, 2012, doi: 10.1021/ci3001277.

[69] J. H. Voigt, B. Bienfait, S. Wang, and M. C. Nicklaus, "Comparison of the NCI Open

Database with Seven Large Chemical Structural Databases," *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 3, pp. 702–712, 2001, doi: 10.1021/ci000150t.

[70] R. Visini, M. Awale, and J. L. Reymond, "Fragment Database FDB-17," *J. Chem. Inf. Model.*, vol. 57, no. 4, pp. 700–709, 2017, doi: 10.1021/acs.jcim.7b00020.

[71] J. Ahmed *et al.*, "FragmentStore-A comprehensive database of fragments linking metabolites, toxic molecules and drugs," *Nucleic Acids Res.*, vol. 39, no. SUPPL. 1, pp. 1049–1054, 2011, doi: 10.1093/nar/gkq969.

[72] E. E. Bolton, S. Kim, and S. H. Bryant, "PubChem3D: Conformer generation," *J. Cheminform.*, vol. 3, no. 1, p. 4, 2011, doi: 10.1186/1758-2946-3-4.

[73] A. Cherkasov *et al.*, "QSAR modeling: Where have you been? Where are you going to?," *J. Med. Chem.*, vol. 57, no. 12, pp. 4977–5010, 2014, doi: 10.1021/jm4004285.

[74] J. P. A. Martins and M. M. C. Ferreira, "QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR," *Quim. Nova*, vol. 36, no. 4, pp. 554–560, 2013, doi: 10.1590/s0100-40422013000400013.

[75] P. Gramatica, "Principles of QSAR Modeling," *Int. J. Quant. Struct. Relationships*, vol. 5, no. 3, pp. 1–37, 2020, doi: 10.4018/ijqspr.20200701.oa1.

[76] J. Leszczynski, A. Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopoulos, H. Reis, and M. K. Shukla, "Handbook of computational chemistry," *Handb. Comput. Chem.*, no. January, pp. 1–2381, 2017, doi: 10.1007/978-3-319-27282-5.

[77] L. P. Hammett, "The effect of structure upon the reactions of organic compounds. temperature and solvent influences," *J. Chem. Phys.*, vol. 4, no. 9, pp. 613–617, 1936, doi: 10.1063/1.1749914.

[78] C. Hansch, "Method Activity," vol. 5, no. 4, pp. 1616–1626, 1963.

[79] J. W. Wilson, S. M. Free, and J. W. Wilson, "eugene," vol. 7, no. 4, 1964.

[80] A. Tropsha, "Best Practices for QSAR Model Development , Validation , and Exploitation," pp. 476–488, 2010, doi: 10.1002/minf.201000061.

[81] D. Fourches, E. Muratov, and A. Tropsha, "Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research," *J. Chem. Inf. Model.*, vol. 50, no. 7, pp. 1189–1204, 2010, doi: 10.1021/ci100176x.

[82] A. Vedani and M. Dobler, "5D-QSAR: The key for simulating induced fit?," *J. Med. Chem.*, vol. 45, no. 11, pp. 2139–2149, 2002, doi: 10.1021/jm011005p.

[83] A. Vedani, M. Dobler, and M. A. Lill, "Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor," *J. Med. Chem.*, vol. 48, no. 11, pp. 3700–3703, 2005, doi: 10.1021/jm050185q.

[84] C. W. Yap, "PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints," *J. Comput. Chem.*, vol. 32, no. 7, pp. 1466–1474, 2011, doi: 10.1002/jcc.

[85] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "DRAGON software: An easy approach to molecular descriptor calculations," *Match*, vol. 56, no. 2, pp. 237–248,

2006.

[86] I. Zamora, T. Oprea, G. Cruciani, M. Pastor, and A. L. Ungell, “Surface descriptors for protein-ligand affinity prediction,” *J. Med. Chem.*, vol. 46, no. 1, pp. 25–33, 2003, doi: 10.1021/jm011051p.

[87] A. Kerber, R. Laue, M. Meringer, and C. Rücker, “Molgen-QSPR, a software package for the study of quantitative structure property relationships,” *Match*, no. 51, pp. 187–204, 2004.

[88] H. Moriwaki, Y. S. Tian, N. Kawashita, and T. Takagi, “Mordred: A molecular descriptor calculator,” *J. Cheminform.*, vol. 10, no. 1, pp. 1–14, 2018, doi: 10.1186/s13321-018-0258-y.

[89] I. B. Grillo, G. A. Urquiza-Carvalho, and G. B. Rocha, “PRIMoRDiA: A Software to Explore Reactivity and Electronic Structure in Large Biomolecules,” *J. Chem. Inf. Model.*, vol. 60, no. 12, pp. 5885–5890, 2020, doi: 10.1021/acs.jcim.0c00655.

[90] A. Dey, “Machine Learning Algorithms: A Review,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016, doi: 10.21275/ART20203995.

[91] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[92] S. Badillo *et al.*, “An Introduction to Machine Learning,” *Clin. Pharmacol. Ther.*, vol. 107, no. 4, pp. 871–885, 2020, doi: 10.1002/cpt.1796.

[93] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised Machine Learning: A Brief Primer,” *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, 2020, doi: 10.1016/j.beth.2020.05.002.

[94] P. Hahn, “Künstliche Intelligenz und Maschinelles Lernen TT - Artificial intelligence and machine learning,” *Handchir Mikrochir Plast Chir*, vol. 51, no. 01, pp. 62–67, 2019.

[95] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–16, 2019, doi: 10.1186/s12911-019-1004-8.

[96] N. A. Danilo Bzdok, Martin Krzywinski, “Machine learning: supervised methods,” *Nat Methods.*, vol. 15, no. 1, pp. 5–6, 2018, doi: 10.1038/nmeth.4551.Machine.

[97] J. D. Maccuish and N. E. Maccuish, “Chemoinformatics applications of cluster analysis,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, no. 1, pp. 34–48, 2014, doi: 10.1002/wcms.1152.

[98] L. B. Akella and D. DeCaprio, “Cheminformatics approaches to analyze diversity in compound screening libraries,” *Curr. Opin. Chem. Biol.*, vol. 14, no. 3, pp. 325–330, 2010, doi: 10.1016/j.cbpa.2010.03.017.

[99] K. A. Marill, “Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression,” *Acad. Emerg. Med.*, vol. 11, no. 1, pp. 94–102, 2004, doi: 10.1197/j.aem.2003.09.006.

[100] M. A. Khanfar and M. O. Taha, “Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR inhibitors,” *J. Chem. Inf. Model.*, vol. 53, no. 10, pp. 2587–2612, 2013, doi: 10.1021/ci4003798.

[101] J. Hert *et al.*, “New methods for ligand-based virtual screening: Use of data fusion and machine learning to enhance the effectiveness of similarity searching,” *J. Chem. Inf. Model.*, vol. 46, no. 2, pp. 462–470, 2006, doi: 10.1021/ci050348j.

[102] T. Huang *et al.*, “MOST: Most-similar ligand based approach to target prediction,” *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–11, 2017, doi: 10.1186/s12859-017-1586-z.

[103] I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, “Neural networks in building QSAR models,” *Methods Mol. Biol.*, vol. 458, no. December, pp. 137–158, 2008, doi: 10.1007/978-1-60327-101-1_8.

[104] W. S. Noble, “What is a support vector machine?,” *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006, doi: 10.1038/nbt1206-1565.

[105] Y. C. Lo, S. E. Rensi, W. Torng, and R. B. Altman, “Machine learning in chemoinformatics and drug discovery,” *Drug Discov. Today*, vol. 23, no. 8, pp. 1538–1546, 2018, doi: 10.1016/j.drudis.2018.05.010.

[106] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003, doi: 10.1021/ci034160g.

[107] U. Muhammad, A. Uzairu, and D. Ebuka Arthur, “Review on: quantitative structure activity relationship (QSAR) modeling,” *J. Anal. Pharm. Res.*, vol. 7, no. 2, pp. 240–242, 2018, doi: 10.15406/japlr.2018.07.00232.

[108] I. Corso and C. Lorena, “Aplicación de algoritmos de clasificación supervisada usando,” p. 11, 2007.

[109] K. H. Zou, A. J. O’Malley, and and Laura Mauri, “Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models,” *Circulation*, vol. 115, no. 1, pp. 654–657, 2007, doi: doi.org/10.1161/CIRCULATIONAHA.105.594929.

[110] T. C. F. Polo and H. A. Miot, “Aplicações da curva ROC em estudos clínicos e experimentais,” *J. Vasc. Bras.*, vol. 19, pp. 23–26, 2020, doi: 10.1590/1677-5449.200186.

[111] S. Gill, A. Christopher, V. Gupta, and P. Bansal, “Emerging role of bioinformatics tools and software in evolution of clinical research,” *Perspect. Clin. Res.*, vol. 7, no. 3, p. 115, 2016, doi: 10.4103/2229-3485.184782.

[112] A. C. Mafud, L. G. Ferreira, Y. P. Mascarenhas, A. D. Andricopulo, and J. de Moraes, “Discovery of Novel Antischistosomal Agents by Molecular Modeling Approaches,” *Trends Parasitol.*, vol. 32, no. 11, pp. 874–886, 2016, doi: 10.1016/j.pt.2016.08.002.

[113] A. J. Trott, O.; Olson, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.,” *J. Comput.*

Chem., vol. 31, no. 2, pp. 455–461, 2010, doi: 10.1002/jcc.21334.

[114] J. de Ruyck, G. Brysbaert, R. Blossey, and M. F. Lensink, "Molecular docking as a popular tool in drug design, an in silico travel," *Adv. Appl. Bioinforma. Chem.*, vol. 9, no. 1, pp. 1–11, 2016, doi: 10.2147/AABC.S105289.

[115] M. Gupta, R. Sharma, and A. Kumar, "Docking techniques in pharmacology: How much promising?," *Comput. Biol. Chem.*, vol. 76, pp. 210–217, 2018, doi: 10.1016/j.compbiolchem.2018.06.005.

[116] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery," *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, 2012, doi: 10.2174/157340911795677602.

[117] D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti, and A. Cavalli, "Dynamic docking: A paradigm shift in computational drug discovery," *Molecules*, vol. 22, no. 11, pp. 1–21, 2017, doi: 10.3390/molecules22112029.

[118] P. H. M. Torres, A. C. R. Sodero, P. Jofily, and F. P. Silva-Jr, "Key Topics in Molecular Docking for Drug Design," *Int. J. Mol. Sci.*, vol. 20, no. 18, p. 4574, 2019, doi: 10.3390/ijms20184574.

[119] S. Saikia and M. Bordoloi, "Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective," *Curr. Drug Targets*, vol. 20, no. 5, pp. 501–521, 2018, doi: 10.2174/1389450119666181022153016.

[120] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A Fast Flexible Docking Method using an Incremental Construction Algorithm," pp. 470–489, 1996.

[121] T. J. A. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz, "DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases," *J. Comput. Aided. Mol. Des.*, vol. 15, no. 5, pp. 411–428, 2001, doi: 10.1023/A:1011115820450.

[122] M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, "FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure," *J. Comput. Aided. Mol. Des.*, vol. 8, no. 2, pp. 153–174, 1994, doi: 10.1007/BF00119865.

[123] M. McGann, "FRED pose prediction and virtual screening accuracy," *J. Chem. Inf. Model.*, vol. 51, no. 3, pp. 578–596, 2011, doi: 10.1021/ci100436p.

[124] T. A. Halgren *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," *J. Med. Chem.*, vol. 47, no. 7, pp. 1750–1759, 2004, doi: 10.1021/jm030644s.

[125] B. S. P. Bentham Science Publisher, "eHiTS: An Innovative Approach to the Docking and Scoring Function Problems," *Curr. Protein Pept. Sci.*, vol. 7, no. 5, pp. 421–435, 2006, doi: 10.2174/138920306778559412.

[126] D. Abagyan, R.; Totrov, M.; Kuznetsov, "ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation.," *J. Comput. Chem.*, vol. 15, no. 1, pp. 488–506, 1994.

[127] A. J. Morris, G. M.;Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R.K.; Goodsell,

D.S.; Olson, “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility,” *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, 2009, doi: 10.1002/jcc.21256.

[128] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, “Improved protein-ligand docking using GOLD,” *Proteins Struct. Funct. Genet.*, vol. 52, no. 4, pp. 609–623, 2003, doi: 10.1002/prot.10465.

[129] K. B. Santos, I. A. Guedes, A. L. M. Karl, and L. E. Dardenne, “Highly Flexible Ligand Docking: Benchmarking of the DockThor Program on the LEADS-PEP Protein-Peptide Data Set,” *J. Chem. Inf. Model.*, vol. 60, no. 2, pp. 667–683, 2020, doi: 10.1021/acs.jcim.9b00905.

[130] R. Thomsen and M. H. Christensen, “MolDock: A new technique for high-accuracy molecular docking,” *J. Med. Chem.*, vol. 49, no. 11, pp. 3315–3321, 2006, doi: 10.1021/jm051197e.

[131] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: Application of ant colony optimization to structure-based drug design,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4150 LNCS, no. September, pp. 247–258, 2006, doi: 10.1007/11839088_22.

[132] M. K. A Miranker, “Functionality maps of binding sites: a multiple copy simultaneous search method,” *Proteins*, vol. 11, no. 1, pp. 29–34, 1991.

[133] G. Wu, D. H. Robertson, C. L. Brooks, and M. Vieth, “Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMm-based MD docking algorithm,” *J. Comput. Chem.*, vol. 24, no. 13, pp. 1549–1562, 2003, doi: 10.1002/jcc.10306.

[134] J. Li, A. Fu, and L. Zhang, “An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking,” *Interdiscip. Sci. Comput. Life Sci.*, vol. 11, no. 2, pp. 320–328, 2019, doi: 10.1007/s12539-019-00327-w.

[135] X. Ren *et al.*, “Novel Consensus Docking Strategy to Improve Ligand Pose Prediction,” *J. Chem. Inf. Model.*, vol. 58, no. 8, pp. 1662–1668, 2018, doi: 10.1021/acs.jcim.8b00329.

[136] E. Pini *et al.*, “New chromane-based derivatives as inhibitors of mycobacterium tuberculosis salicylate synthase (mbti): Preliminary biological evaluation and molecular modeling studies,” *Molecules*, vol. 23, no. 7, 2018, doi: 10.3390/molecules23071506.

[137] D. D. Li *et al.*, “Consensus scoring model for the molecular docking study of mTOR kinase inhibitor,” *J. Mol. Graph. Model.*, vol. 79, pp. 81–87, 2018, doi: 10.1016/j.jmgm.2017.11.003.

[138] A. T. Onawole, K. O. Sulaiman, R. O. Adegoke, and T. U. Kolapo, “Identification of potential inhibitors against the Zika virus using consensus scoring,” *J. Mol. Graph. Model.*, vol. 73, pp. 54–61, 2017, doi: 10.1016/j.jmgm.2017.01.018.

[139] L. Chaput, J. Martinez-Sanz, E. Quiniou, P. Rigolet, N. Saettel, and L. Mouawad, “VSDC: A method to improve early recognition in virtual screening when limited experimental resources are available,” *J. Cheminform.*, vol. 8, no. 1, pp. 1–18, 2016,

doi: 10.1186/s13321-016-0112-z.

[140] D. R. Houston and M. D. Walkinshaw, “Consensus docking: Improving the reliability of docking in a virtual screening context,” *J. Chem. Inf. Model.*, vol. 53, no. 2, pp. 384–390, 2013, doi: 10.1021/ci300399w.

[141] M. dos Santos Maia, G. C. Soares Rodrigues, A. B. Silva Cavalcanti, L. Scotti, and M. T. Scotti, “Consensus Analyses in Molecular Docking Studies Applied to Medicinal Chemistry,” *Mini-Reviews Med. Chem.*, vol. 20, no. 14, pp. 1322–1340, 2020, doi: 10.2174/1389557520666200204121129.

[142] P. C. Do, E. H. Lee, and L. Le, “Steered Molecular Dynamics Simulation in Rational Drug Design,” *J. Chem. Inf. Model.*, vol. 58, no. 8, pp. 1473–1482, 2018, doi: 10.1021/acs.jcim.8b00261.

[143] A. Pérez, G. Martínez-Rosell, and G. De Fabritiis, “Simulations meet machine learning in structural biology,” *Curr. Opin. Struct. Biol.*, vol. 49, pp. 139–144, 2018, doi: 10.1016/j.sbi.2018.02.004.

[144] S. A. Hollingsworth and R. O. Dror, “Molecular Dynamics Simulation for All,” *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018, doi: 10.1016/j.neuron.2018.08.011.

[145] M. Bermudez, J. Mortier, C. Rakers, D. Sydow, and G. Wolber, “More than a look into a crystal ball: protein structure elucidation guided by molecular dynamics simulations,” *Drug Discov. Today*, vol. 21, no. 11, pp. 1799–1805, 2016, doi: 10.1016/j.drudis.2016.07.001.

[146] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of Molecular Dynamics and Related Methods in Drug Discovery,” *J. Med. Chem.*, vol. 59, no. 9, pp. 4035–4061, 2016, doi: 10.1021/acs.jmedchem.5b01684.

[147] X. Liu, D. Shi, S. Zhou, H. Liu, H. Liu, and X. Yao, “Molecular dynamics simulations and novel drug discovery,” *Expert Opin. Drug Discov.*, vol. 13, no. 1, pp. 23–37, 2018, doi: 10.1080/17460441.2018.1403419.

[148] M. C. C. and V. Daggett, “Insights from molecular dynamics simulations for computational protein design,” *Mol Syst Des Eng*, vol. 2, no. 1, pp. 9–33, 2017, doi: 10.1039/C6ME00083E.

[149] S. H. Chong, P. Chatterjee, and S. Ham, “Computer Simulations of Intrinsically Disordered Proteins,” *Annu. Rev. Phys. Chem.*, vol. 68, pp. 117–134, 2017, doi: 10.1146/annurev-physchem-052516-050843.

[150] A. Ganesan, M. L. Coote, and K. Barakat, “Molecular dynamics-driven drug discovery: leaping forward with confidence,” *Drug Discov. Today*, vol. 22, no. 2, pp. 249–269, 2017, doi: 10.1016/j.drudis.2016.11.001.

[151] M. Rouhani, F. Khodabakhsh, D. Norouzian, R. A. Cohan, and V. Valizadeh, “Molecular dynamics simulation for rational protein engineering: Present and future prospectus,” *J. Mol. Graph. Model.*, vol. 84, pp. 43–53, 2018, doi: 10.1016/j.jmgm.2018.06.009.

[152] J. Wang, P. Cieplak, and P. A. Kollman, “How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic

and Biological Molecules?," *J. Comput. Chem.*, vol. 21, no. 12, pp. 1049–1074, 2000, doi: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F.

[153] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins Struct. Funct. Genet.*, vol. 65, no. 3, pp. 712–725, 2006, doi: 10.1002/prot.21123.

[154] K. Lindorff-Larsen *et al.*, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins Struct. Funct. Bioinforma.*, vol. 78, no. 8, pp. 1950–1958, 2010, doi: 10.1002/prot.22711.

[155] P. S. Nerenberg and T. Head-Gordon, "Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides," *J. Chem. Theory Comput.*, vol. 7, no. 4, pp. 1220–1230, 2011, doi: 10.1021/ct2000183.

[156] D. W. Li and R. Brüschweiler, "NMR-based protein potentials," *Angew. Chemie - Int. Ed.*, vol. 49, no. 38, pp. 6778–6780, 2010, doi: 10.1002/anie.201001898.

[157] Y. Duan *et al.*, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations," *J. Comput. Chem.*, vol. 24, no. 16, pp. 1999–2012, 2003, doi: 10.1002/jcc.10349.

[158] R. B. Best and J. Mittal, "Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse," *J. Phys. Chem. B*, vol. 114, no. 46, pp. 14916–14923, 2010, doi: 10.1021/jp108618d.

[159] A. D. MacKerell *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998, doi: 10.1021/jp973084f.

[160] A. D. Mackerell, M. Feig, and C. L. Brooks, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation," *J. Comput. Chem.*, vol. 25, no. 11, pp. 1400–1415, 2004, doi: 10.1002/jcc.20065.

[161] M. Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal and and A. D. M. J. Feig, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and sidechain χ1 and χ2 dihedral angles," *J Chem Theory Comput.*, vol. 8, no. 9, pp. 3257–3273, 2012, doi: 10.1021/ct300400x.Optimization.

[162] L. Pol-Fachin, C. L. Fernandes, and H. Verli, "GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations," *Carbohydr. Res.*, vol. 344, no. 4, pp. 491–500, 2009, doi: 10.1016/j.carres.2008.12.025.

[163] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1656–1676, 2004, doi: 10.1002/jcc.20090.

[164] N. Schmid *et al.*, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *Eur. Biophys. J.*, vol. 40, no. 7, pp. 843–856, 2011, doi: 10.1007/s00249-

011-0700-9.

[165] G. A. Kaminski, R. A. Friesner, J. Tirado-rives, and W. L. Jorgensen, “Comparison with Accurate Quantum Chemical Calculations on Peptides †,” *Quantum*, vol. 2, pp. 6474–6487, 2001.

[166] M. J. Abraham *et al.*, “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1–2, pp. 19–25, 2015, doi: 10.1016/j.softx.2015.06.001.

[167] J. R. Biller, H. Elajaili, V. Meyer, G. M. Rosen, S. S. Eaton, and G. R. Eaton, “The Amber biomolecular simulation programs.,” *J. Magn. Reson.*, vol. 236, no. 16, pp. 47–56, 2013.

[168] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *Parallel Sci. Eng. Appl. Charm++ Approach*, vol. 26, no. 16, pp. 60–76, 2016, doi: 10.1201/b16251-15.

[169] B. R. Brooks *et al.*, “CHARMM: The biomolecular simulation program,” *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009, doi: 10.1002/jcc.21287.

[170] A. D. Astuti, R. Refianti, and A. B. Mutiara, “Molecular Dynamics Simulation on Protein Using Gromacs,” *J. Comput. Sci.*, vol. 9, no. 2, pp. 16–20, 2011.

[171] B. W. and A. Sali, “Comparative Protein Structure Modeling Using MODELLER,” *Curr Protoc Bioinforma.*, vol. 54, no. 5, pp. 1–5, 2017.

[172] Samuel Genheden & Ulf Ryde, “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities,” *Expert Opin. Drug Discov.*, vol. 10, no. 5, pp. 449–461, 2015, doi: 10.1517/17460441.2015.1032936.

[173] R. Kumari, R. Kumar, and A. Lynn, “G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations,” *J. Chem. Inf. Model.*, vol. 54, no. 7, pp. 1951–1962, 2014, doi: 10.1021/ci500020m.

[174] L. C. Martins, P. H. M. Torres, R. B. de Oliveira, P. G. Pascutti, E. A. Cino, and R. S. Ferreira, “Investigation of the binding mode of a novel cruzain inhibitor by docking, molecular dynamics, ab initio and MM/PBSA calculations,” *J. Comput. Aided. Mol. Des.*, vol. 32, no. 5, pp. 591–605, 2018, doi: 10.1007/s10822-018-0112-3.

[175] Z. Dolatkhah, S. Javanshir, A. S. Sadr, J. Hosseini, and S. Sardari, “Synthesis, Molecular Docking, Molecular Dynamics Studies, and Biological Evaluation of 4H-Chromone-1,2,3,4-tetrahydropyrimidine-5-carboxylate Derivatives as Potential Antileukemic Agents,” *J. Chem. Inf. Model.*, vol. 57, no. 6, pp. 1246–1257, 2017, doi: 10.1021/acs.jcim.6b00138.

[176] J. A. Arnott and S. L. Planey, “The influence of lipophilicity in drug discovery and design,” *Expert Opin. Drug Discov.*, vol. 7, no. 10, pp. 863–875, 2012, doi: 10.1517/17460441.2012.714363.

[177] Y. C. Martin, “A bioavailability score,” *J. Med. Chem.*, vol. 48, no. 9, pp. 3164–3170, 2005, doi: 10.1021/jm0492002.

[178] C. F. Lipinski, V. G. Maltarollo, P. R. Oliveira, A. B. F. da Silva, and K. M. Honorio, “Advances and Perspectives in Applying Deep Learning for Drug Design and

Discovery," *Front. Robot. AI*, vol. 6, no. November, pp. 1–6, 2019, doi: 10.3389/frobt.2019.00108.

[179] A. Daina, O. Michielin, and V. Zoete, "ILOGP: A simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach," *J. Chem. Inf. Model.*, vol. 54, no. 12, pp. 3284–3301, 2014, doi: 10.1021/ci500467k.

[180] A. Daina and V. Zoete, "A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules," *ChemMedChem*, pp. 1117–1121, 2016, doi: 10.1002/cmdc.201600182.

[181] A. Daina, O. Michielin, and V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–13, 2017, doi: 10.1038/srep42717.

[182] U. M. Zanger and M. Schwab, "Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation," *Pharmacol. Ther.*, vol. 138, no. 1, pp. 103–141, 2013, doi: 10.1016/j.pharmthera.2012.12.007.

[183] Z. Zhang and W. Tang, "Drug metabolism in drug discovery and development," *Acta Pharm. Sin. B*, vol. 8, no. 5, pp. 721–732, 2018, doi: 10.1016/j.apsb.2018.04.003.

[184] B. Testa, "Metabolism in Drug Development," in *Drug Metabolism Prediction*, 2014, pp. 1–26.

[185] A. Daina, O. Michielin, and V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–13, 2017, doi: 10.1038/srep42717.

[186] K. M. Raghava and P. K. Lakshmi, "Overview of P-glycoprotein inhibitors: A rational outlook," *Brazilian J. Pharm. Sci.*, vol. 48, no. 3, pp. 353–367, 2012, doi: 10.1590/S1984-82502012000300002.

[187] E. Choong, M. Dobrinas, P. A. Carrupt, and C. B. Eap, "The permeability P-glycoprotein: A focus on enantioselectivity and brain distribution," *Expert Opin. Drug Metab. Toxicol.*, vol. 6, no. 8, pp. 953–965, 2010, doi: 10.1517/17425251003789394.

[188] R. Parthasarathi and A. Dhawan, *In Silico Approaches for Predictive Toxicology*. Elsevier Inc., 2018.

[189] M. A. Dorato and L. A. Buckley, "Toxicology in the drug discovery and development process.," *Curr. Protoc. Pharmacol.*, vol. Chapter 10, pp. 1–35, 2006, doi: 10.1002/0471141755.ph1003s32.

[190] N. M. Luscombe, D. Greenbaum, and M. Gerstein, "What is bioinformatics? A proposed definition and overview of the field," *Methods Inf. Med.*, vol. 40, no. 4, pp. 346–358, 2001, doi: 10.1055/s-0038-1634431.

[191] X. Xia, "Bioinformatics and Drug Discovery," *Curr. Top. Med. Chem.*, vol. 17, no. 15, pp. 1709–1726, 2017, doi: 10.2174/1568026617666161116143440.

[192] M. Arauzo-Bravo and S. Ahmad, “Protein Sequence and Structure Databases: A Review,” *Curr. Anal. Chem.*, vol. 1, no. 3, pp. 355–371, 2005, doi: 10.2174/157341105774573974.

[193] R. J. Owens, “Structural proteomics: High-throughput methods: Second edition,” *Struct. Proteomics High-Throughput Methods Second Ed.*, pp. 1–371, 2014, doi: 10.1007/978-1-4939-2230-7.

[194] C. H. pd. Sers/tWu, C. Arighi, and K. Ross, *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*, vol. 1558, no. 1558. 2017.

[195] M. D. Brice, J. R. Rodgers, and O. Kennard, “The Protein Data Bank,” vol. 324, pp. 319–324, 1977.

[196] J. L. Moreland, A. Gramada, O. V. Buzko, Q. Zhang, and P. E. Bourne, “The Molecular Biology Toolkit (MBT): A modular platform fro developing molecular visualization applications,” *BMC Bioinformatics*, vol. 6, pp. 1–7, 2005, doi: 10.1186/1471-2105-6-21.

[197] D. Piovesan *et al.*, “MobiDB: Intrinsically disordered proteins in 2021,” *Nucleic Acids Res.*, vol. 49, no. D1, pp. D361–D367, 2021, doi: 10.1093/nar/gkaa1058.

[198] U. Pieper *et al.*, “ModBase,a database of annotated comparative protein structure models,and associated resources,” *Nucleic Acids Res.*, vol. 39, no. SUPPL. 1, pp. 465–474, 2011, doi: 10.1093/nar/gkq1091.

[199] A. Gutmanas *et al.*, “PDBe: Protein data bank in Europe,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 285–291, 2014, doi: 10.1093/nar/gkt1180.

[200] A. R. Kinjo *et al.*, “Protein Data Bank Japan (PDBj): Updated user interfaces, resource description framework, analysis tools for large structures,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D282–D288, 2017, doi: 10.1093/nar/gkw962.

[201] R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, and J. M. Thornton, “PDBsum: Structural summaries of PDB entries,” *Protein Sci.*, vol. 27, no. 1, pp. 129–134, 2018, doi: 10.1002/pro.3289.

[202] K. Arnold *et al.*, “The Protein Model Portal,” *J. Struct. Funct. Genomics*, vol. 10, no. 1, pp. 1–8, 2009, doi: 10.1007/s10969-008-9048-5.

[203] S. K. Burley *et al.*, “RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences,” *Nucleic Acids Res.*, vol. 49, no. 1, pp. D437–D451, 2021, doi: 10.1093/nar/gkaa1038.

[204] A. M. Amaral, M. S. Reis, and F. R. Silva, “O programa BLAST : guia prático de utilização,” *Embrapa Recur. Genéticos e Biotecnol.*, p. 24, 2007.

[205] W. J. S. Diniz and F. Canduri, “Bioinformatics: An overview and its applications,” *Genet. Mol. Res.*, vol. 16, no. 1, pp. 1–21, 2017, doi: 10.4238/gmr16019645.

[206] Hugo Verli (Org.), *Bioinformática da Biologia à Flexibilidade Molecular*. 2014.

[207] M. Dorn, M. B. E Silva, L. S. Buriol, and L. C. Lamb, “Three-dimensional protein structure prediction: Methods and computational strategies,” *Comput. Biol. Chem.*, vol. 53, no. PB, pp. 251–276, 2014, doi: 10.1016/j.compbiolchem.2014.10.001.

[208] D. OSGUTHORPE, “Ab initio protein folding,” *Curr. Opin. Struct. Biol*, vol. 10, no. 2, pp. 146–152, 2000, doi: doi.org/10.1016/S0959-440X(00)00067-1.

[209] V. V. A. Steven M. Singer#, Marc Y. Fink, “Protein structure prediction,” *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2019, doi: 10.1142/S021797921840009X.Protein.

[210] Y. Ihm, “A threading approach to protein structure prediction: studies on TNF-like molecules, Rev proteins, and protein kinases Recommended Citation,” 2004.

[211] V. K. Vyas, R. D. Ukawala, M. Ghate, and C. Chintha, “Homology modeling a fast tool for drug discovery: Current perspectives,” *Indian J. Pharm. Sci.*, vol. 74, no. 1, pp. 1–17, 2012, doi: 10.4103/0250-474X.102537.

[212] C. N. Cavasotto and S. S. Phatak, “Homology modeling in drug discovery: current trends and applications,” *Drug Discov. Today*, vol. 14, no. 13–14, pp. 676–683, 2009, doi: 10.1016/j.drudis.2009.04.006.

[213] M. T. Muhammed and E. Aki-Yalcin, “Homology modeling in drug discovery: Overview, current applications, and future perspectives,” *Chem. Biol. Drug Des.*, vol. 93, no. 1, pp. 12–20, 2019, doi: 10.1111/cbdd.13388.

[214] Y. Zhang, “I-TASSER server for protein 3D structure prediction,” *BMC Bioinformatics*, vol. 9, pp. 1–8, 2008, doi: 10.1186/1471-2105-9-40.

[215] A. Waterhouse *et al.*, “SWISS-MODEL: Homology modelling of protein structures and complexes,” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, 2018, doi: 10.1093/nar/gky427.

[216] J. K. X. Maier and P. Labute, “Assessment of fully automated antibody homology modeling protocols in molecular operating environment,” *Proteins Struct. Funct. Bioinforma.*, vol. 82, no. 8, pp. 1599–1610, 2014, doi: 10.1002/prot.24576.

[217] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, “The Phyre2 web portal for protein modeling, prediction and analysis,” *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, 2016, doi: 10.1038/nprot.2015-053.

[218] D. E. Kim, D. Chivian, and D. Baker, “Protein structure prediction and analysis using the Robetta server,” *Nucleic Acids Res.*, vol. 32, no. WEB SERVER ISS., pp. 526–531, 2004, doi: 10.1093/nar/gkh468.

[219] J. Haas *et al.*, “The protein model portal - A comprehensive resource for protein structure and model information,” *Database*, vol. 2013, no. 8, pp. 1–8, 2013, doi: 10.1093/database/bat031.

[220] R. A. T Cardozo , M Totrov, “Homology modeling by the ICM method,” *Proteins*, vol. 23, no. 3, pp. 403–414, 1995.

[221] P. Rathi Suganya, K. Sudevan, S. Kalva, and L. M. Saleena, “Homology modeling for human ADAM12 using prime, I-tasser and easymodeller,” *Int. J. Pharm. Pharm. Sci.*, vol. 6, no. SUPPL. 2, pp. 782–786, 2014.

[222] A. A. C. & R. L. D. J. Qiang Wang, "SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling," *Nat. Protoc. Vol.*, vol. 3, no. 1, pp. 1832–1847, 2008.

[223] B. K. Ho, A. Thomas, and R. Brasseur, "Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α-helix," *Protein Sci.*, vol. 12, no. 11, pp. 2508–2522, 2009, doi: 10.1110/ps.03235203.

[224] B. K. Ho and R. Brasseur, "The Ramachandran plots of glycine and pre-proline," *BMC Struct. Biol.*, vol. 5, pp. 1–11, 2005, doi: 10.1186/1472-6807-5-14.

[225] S. A. Hollingsworth and P. A. Karplus, "A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins," *Biomol. Concepts*, vol. 1, no. 3–4, pp. 271–283, 2010, doi: 10.1515/bmc.2010.022.

[226] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *J. Mol. Biol.*, vol. 7, no. 1, pp. 95–99, 1963, doi: 10.1016/S0022-2836(63)80023-6.

[227] D. Eisenberg, R. Lüthy, and J. U. Bowie, "VERIFY3D: Assessment of protein models with three-dimensional profiles," *Methods Enzymol.*, vol. 277, no. April, pp. 396–404, 1997, doi: 10.1016/S0076-6879(97)77022-8.

[228] D. E. JAMES U. BOWIE, ROLAND LtCY, "A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Stucture," *SCIENCE*, vol. 253, no. 58, 1991.

[229] R. O. Cosentino and F. Agüero, "Genetic profiling of the isoprenoid and sterol biosynthesis pathway genes of Trypanosoma cruzi," *PLoS One*, vol. 9, no. 5, 2014, doi: 10.1371/journal.pone.0096762.

[230] B. S. Shastry, "Pharmacogenetics and the concept of individualized medicine," *Pharmacogenomics J.*, vol. 6, no. 1, pp. 16–21, 2006, doi: 10.1038/sj.tpj.6500338.

[231] J. L. A. John F Carlquist, "Pharmacogenetic mechanisms underlying unanticipated drug responses," *Discov Med .*, vol. 11, no. 60, pp. 469-478., 2011.

[232] J. J.G. Marin, O. Briz, M. J. Monte, A. G. Blazquez, and R. I.R. Macias, "Genetic Variants in Genes Involved in Mechanisms of Chemoresistance to Anticancer Drugs," *Curr. Cancer Drug Targets*, vol. 12, no. 4, pp. 402–438, 2012, doi: 10.2174/156800912800190875.

[233] A. J. Brookes, "The essence of SNPs," *Gene*, vol. 234, no. 2, pp. 177–186, 1999, doi: 10.1016/S0378-1119(99)00219-X.

[234] N. A. Mahmoud, E. I. Mohammed, S. O. Elshafei, A. H. Elsadig, R. A. Omer, and S. B. Mohamed, "Bioinformatics Analysis of Single Nucleotide Polymorphism in Multiple Drug Resistance Protein 1 (ABCC1) in Bos taurus," vol. 7, no. 1, pp. 9–24, 2017, doi: 10.5923/j.bioinformatics.20170701.02.

[235] R. Thomas A. Sellers, Yifan Huang, Julie Cunningham, Ellen L. Goode, M. Sutphen, Robert A. Vierkant, Linda E. Kelemen, Zachary S. Fredericksen3, J. Liebow, V. Shane Pankratz, Lynn C. Hartmann, Jeff Myer, Edwin S. Iversen Jr., and and C. P. M. Schildkraut, "Association of Single Nucleotide Polymorphisms in Glycosylation Genes

with Risk of Epithelial Ovarian Cancer," *Cancer Epidemiol Biomarkers Prev*, vol. 17, no. 2, pp. 397–404, 2008, doi: 10.1158/1055-9965.EPI-07-0565.Association.

[236] Y. Yang, W. Wang, G. Liu, Y. Yu, and M. Liao, "Association of single nucleotide polymorphism rs3803662 with the risk of breast cancer," *Sci. Rep.*, vol. 6, pp. 1–6, 2016, doi: 10.1038/srep29008.

[237] S. Shimizu *et al.*, "Association of single nucleotide polymorphisms in the NRF2 promoter with vascular stiffness with aging," *PLoS One*, vol. 15, no. 8 August, pp. 1–17, 2020, doi: 10.1371/journal.pone.0236834.

[238] C. J. Willer *et al.*, "Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes," *Diabetes*, vol. 56, no. 1, pp. 256–264, 2007, doi: 10.2337/db06-0461.

[239] C. L. L. Valentim *et al.*, "Specific drug action in Schistosome parasites," *Science*, vol. 342, no. 6164, pp. 1385–1389, 2014, doi: 10.1126/science.1243106.Genetic.

Capítulo 2

Neste capítulo foi realizada a triagem virtual de lignanas obtidas do banco de dados ChEMBL. Inicialmente, as lignanas foram submetidas à vários parâmetros farmacocinéticos para obter os compostos com melhores perfis. Em seguida, uma série de ferramentas computacionais de Quimio e Bioinformática foram usadas para classificar e verificar as estruturas químicas potenciais contra espécies de *Leishmania major* e *Leishmania braziliensis*.

As lignanas com melhores perfis farmacocinéticos foram submetidas à predição de atividade biológica através de dois modelos construídos a partir de bancos de dados com atividade biológica contra *L. major* e *L. braziliensis*. Para isso, foram calculados os descritores moleculares no software Dragon, usando o banco de dados do ChEMBL dos compostos com atividade biológica contra as espécies em estudo e as lignanas que passaram nos filtros ADMET. Em seguida, foram utilizados como dados de entrada para a construção dos modelos e a predição de atividade.

O alinhamento é a primeira etapa para a construção de modelos tridimensionais de proteínas. Essa ferramenta é necessária para a investigação de aminoácidos compartilhados entre diferentes espécies e por isso contribui para a análise de sítio ativo. Essa ferramenta foi utilizada para construir modelos de proteínas de *L. major* e, especialmente de *L. braziliensis*, devido a pouca quantidade de estruturas tridimensionais disponíveis experimentalmente.

Uma análise consenso utilizando os dados de predição de atividade biológica e docking molecular foi usado para identificar as lignanas potenciais e *multi-target*. Esse tipo de análise é mais robusta e aumenta a probabilidade de encontrar compostos potenciais contra as espécies estudadas.

Uma análise de polimorfismos de nucleotídeos únicos com alta frequência alélica e de aminoácidos não sinônimos próximo ao sítio ativo das enzimas estudadas foi realizada para avaliar a probabilidade de resistência às lignanas selecionadas neste trabalho. Após os resultados, os complexos com as lignanas com melhores perfis farmacocinéticos, farmacodinâmicos e ativos contra as espécies investidadas foram submetidos à cálculos de dinâmica molecular para avaliação da flexibilidade das proteínas e possíveis alterações nas interações. Por fim, dentre essas lignanas, quatro foram isoladas e testadas contra formas promastigotas de *L. major* e *L. braziliensis* para validação da triagem virtual.

Virtual screening and the in vitro assessment of the antileishmanial activity of lignans

Mayara dos Santos Maia [1], Joanda Paolla Raimundo e Silva [2], Thaís Amanda de Lima Nunes [3], Julyanne Maria Saraiva de Sousa [3], Gabriela Cristina Soares Rodrigues [1], Alex France Messias Monteiro [1], Josean Fechine Tavares [2], Klinger Antonio da Franca Rodrigues [3], Francisco Jaime B. Mendonça-Junior [4], Luciana Scotti [1] and Marcus Tullius Scotti [1,*]

[1] Laboratory of Cheminformatics, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, 58051-900, João Pessoa-PB, Brazil; mayarasmaia@hotmail.com (M.d.S.M.); gaby.ecologia@gnail.com (G.C.S.R.); alexfrancem@ltf.ufpb.br (A.F.M.M.); luciana.scotti@gmail.com (L.S.)

[2] Multi-User Characterization and Analysis Laboratory, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, 58051-900, João Pessoa-PB, Brazil; joanda@ltf.ufpb.br (J.P.R.e.S.); josean@ltf.ufpb.br (J.F.T.)

[3] Infectious Diseases Laboratory, Federal University of Parnaíba Delta, São Benedito, 64202-020, Parnaíba-PI, Brazil; thaisaln13@gmail.com; (T.A.d.L.N.); jully.yanne@gmail.com (J.M.S.d.S.); klinger.antonio@gmail.com (K.A.d.F.R.)

[4] Laboratory of Synthesis and Drug Delivery, State University of Paraíba, 58071-160, João Pessoa-PB, Brazil; franciscojbmendonca@yahoo.com.br

***** Correspondence: mtscotti@gmail.com

Received: 18 April 2020; Accepted: 7 May 2020; Published: 12 May 2020.

Artigo publicado na revista: **Molecules**, 2020, 25(10), 2281.

DOI: 10.3390/molecules25102281

Fator de impacto: 4.587

ABSTRACT: Leishmaniasis is endemic in at least 98 countries. Due to the high toxicity and resistance associated with the drugs, we chose lignans as an alternative, due to their favorable properties of absorption, distribution, metabolism, excretion, and toxicity (ADMET). To investigate their leishmanicidal potential, the biological activities of a set of 160 lignans were predicted using predictive models that were built using data for *Leishmania major* and *L. (Viannia) braziliensis*. A combined analysis, based on ligand and structure, and several other computational approaches were used. The results showed that the combined analysis was able to select 11 lignans with potential activity against *L. major* and 21 lignans against *L. braziliensis*, with multitargeting effects and low or no toxicity. Of these compounds, four were isolated from the species *Justicia aequilabris* (Nees) Lindau. All of the identified compounds were able to inhibit the growth of *L. braziliensis* promastigotes, with the most active compound, (**159**) epipinoresinol-4-*O*-β-D-glucopyranoside, presenting an IC_{50} value of 5.39 µM and IC_{50} value of 36.51 µM for *L. major*. Our findings indicated the potential of computer-aided drug design and development and demonstrated that lignans represent promising prototype compounds for the development of multitarget drugs against leishmaniasis.

Keywords: leishmaniasis; lignan; virtual screening; molecular docking; computer-aided drug design

1. INTRODUCTION

Leishmaniasis refers to a collection of diseases, caused by intracellular parasites of the *Leishmania* genus [1]. Transmission can occur in humans through the bite of Diptera, in the Psychodida family, which are hematophagous insects in the genera *Phlebotomus* or *Lutzomyia* [2]. Leishmaniasis is endemic in at least 98 countries [3] and is considered to be a neglected disease because it affects primarily low-income populations but remains a relevant public health problem, with considerable incidence and prevalence rates in Brazil [4]. The World Health Organization considers leishmaniasis to be one of the six most neglected tropical diseases, affecting 2 million people annually, worldwide [5]. More than 20,000 deaths are estimated to occur annually, due to disease complications [6].

According to Akhoundi et al., 21 species of the genus Leishmania are known to be pathogenic to humans, including *L. major*, *L. tropica*, *L. aethiopica*, *L. donovani*, *L. infantum*, *L. mexicana*, *L. amazonensis*, *L. braziliensis*, *L. peruviana*, *L. guyanensis*, *L. panamemsis*, *L shawi*, *L. lainsoni*, *L. lindenbergi*, *L. martiniquensis*, *L. siamensis*, and *L. colombiensis*. [7]. Leishmaniasis can be categorized as tegumental and visceral leishmaniasis, and both can be further subdivided into other clinical forms of infection, defined by the location of the parasite in infected tissues and by the species, including cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL), visceral or kala azalis leishmaniasis (VL), and post-

kala azal dermal leishmaniasis (PKDL) [8]. The most common form is CL, and more than 90% of all CL cases are distributed across three primary regions: (i) Afghanistan, Iran, Saudi Arabia, and Syria; ii) Algeria and Tunisia; and (iii) Brazil and Peru [9].

CL symptoms present as one or more lesions, which are often painless, and purulence is uncommon. Lesions are typically located in exposed areas, such as the face and extremities, with ulcers and nodule/plaque appearances being common, and transmission occurs through the bites of mosquitos infected with the parasite [10]. Although not fatal, multiple lesions tend to leave permanent scars that can lead to social stigmatization.

In the Americas, the primary etiological agent of CL is *L. braziliensis* [11], which generally encompasses severe clinical forms associated with skin, mucosal, mucocutaneous, and subcutaneous nodular lesions [12,13], and large genetic polymorphisms can cause visible deformities in the host, in addition to psychological, social, and economic impacts [14]. CL is a chronic inflammatory disease that is widely distributed in Brazil and primarily caused by *L. braziliensis* [15].

In India, the Middle East, Central Asia, and North and West Africa, CL is primarily caused by the species *L. major* [12,16], causing mild to severe skin disorders that can result in disfigurement if left untreated [17]. In the Middle Eastern region and Israel, CL is frequently caused in humans by infection through wild mammal reservoirs, with *Phlebotomus papatasi* being the primary vector for *L. major* [16]. Although *Leishmania* parasites affect millions of people, in several countries, around the world, no human vaccine is currently available for the treatment of CL caused by *L. major,* and the available treatments are expensive and have toxic side-effects. Thus, the identification of new therapeutic options against the disease remains necessary [18].

Current treatments for leishmaniasis are based on chemotherapeutic drugs, which are often inefficient or harmful, due to the development of drug resistance and side effects, associated with the high toxicity index. In addition, no vaccines are currently available, although some vaccine candidates for the treatment of these diseases are currently in the pre-clinical and clinical testing phases [19,20].

Because *L. major* and *L. braziliensis* cause lesions in the host that can be disseminated to other sites and the exacerbated production of cytokines and chemokines that cause oxidative stress, trigger the amplification of the inflammatory response [21], we chose liganans and neolignans because they have properties favorable to drug development. Factors in the host such as immunosuppressant, malnutrition, and co-infection or genetic and environmental factors are factors that aggravate the disease. In this sense, we chose lignans and neolignans to

investigate antileishmania activity in enzymes important for the survival and proliferation of parasites, reducing injuries and decrease the inflammatory response [22]. In addition, lignans are known for their anti-inflammatory and antioxidant activity, which could minimize the effects of the inflammatory response. In addition, a study by Pilkington [23], evaluating the absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles of lignans found that more than 75% of lignans met all the requirements for drug-likeness. The study concluded that lignans show a high level of drug similarity.

Computational tools can contribute greatly to database creation, by predicting protein functions, modeling protein structures, simulating metabolic pathway kinetics, predicting biological activities, predicting toxicity, and predicting the affinities and flexibilities between receptors and ligands, which can facilitate the development and identification of drugs with the potential to treat various diseases and promote the development of efficacious drugs with reduced toxicity [24,25].

Therefore, this study aimed to use virtual screening and experimental validation to identify lignans with leishmanicidal potential, low toxicity, and selective activity against several Leishmania targets.

2. RESULTS

2.1 Prediction of ADMET Properties

Various predictive parameters were determined for a set of 160 lignans and neolignans, to identify compounds with the best ADMET profiles for further examination using other methodologies. The results showed that among the 160 lignans and neolignanas, only 34 failed the Lipinski rule. Because the application of the Lipinski rule did not decisively filter the molecules, we used additional methodologies to select those compounds with the best profiles.

During the analysis of lipophilicity and water solubility, 148 compounds (92%) obtained good results, presenting consensus log *p* values below <4.15 and/or at least two descriptors with the classification “Low solubility” (Annex I: Table S1). Then, the 148 compounds were submitted to pharmacokinetic analyzes. The results showed that 42 lignans (28.3%) had adequate pharmacokinetics (Annex I: Table S1).

Toxicity was assessed for the 42 lignans and we found that 33 (78%) of the 42 compounds with good pharmacokinetic action had low or no predicted risk for the development of mutagenicity, tumorigenesis, negative effects on the reproductive system, or irritability (Annex I: Table S2).

2.2 Quantitative Structure-Activity Relationship (QSAR) Modeling

To perform ligand-based virtual screening, two prediction models were built, using the random forest (RF) algorithm. To construct these models, molecular descriptors were calculated for the bank of molecules with known activity against *L. major* and *L. braziliensis*, obtained through the ChEMBL database. After validating the models, 33 lignans with excellent ADMET profiles were analyzed for leishmanicidal activity, using the prediction models.

The RF models were evaluated for their predictive powers, using the parameters of specificity, sensitivity, accuracy, positive predicted value (PPV), and negative predicted value (NPV), in addition to performance and robustness, such as the receiver operating characteristic (ROC) curve and Mathews correlation coefficient (MCC). Table 1 describes the characteristics of the two models, in terms of predictive power and robustness, and Figures 1 **and 2** show the performances of the models.

Table 1. Summary of parameters corresponding to the results obtained for all models.

Specie	Validation	Specificity	Sensitivity	Accuracy	PPV	NPV	MCC
L. major	Test	0.81	0.82	0.81	0.83	0.79	0.63
	Cross	0.80	0.83	0.82	0.83	0.80	0.63
L. braziliensis	Test	0.83	0.73	0.79	0.76	0.80	0.57
	Cross	0.85	1	0.91	0.83	1	0.87

The cross-validation results demonstrated that the generated models obtained excellent performance results, with an accuracy greater than 76%. The *L. major* ROC curve showed a value greater than 0.89 and the MCC value was 0.63, which indicated that the model demonstrated excellent classification, performance, and robustness rates. The ROC curve for the *L. braziliensis* model was greater than 0.86, with an MCC value equal to 0.87, which were also good results.

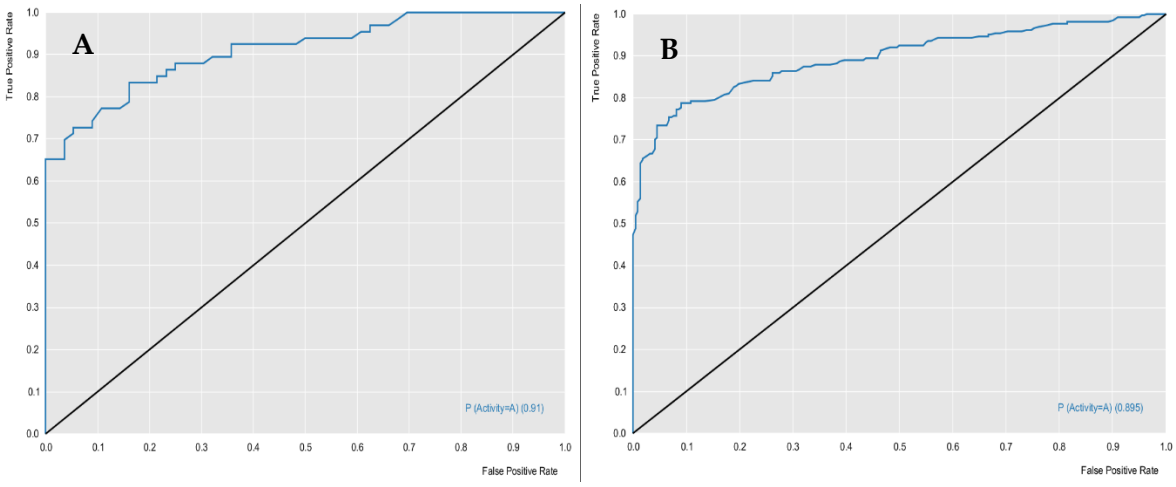


Figure 1. Receiver operating characteristic (ROC) curve generated for the *L. major* random forest (RF) model. (**A**) Test and (**B**) cross-validation.

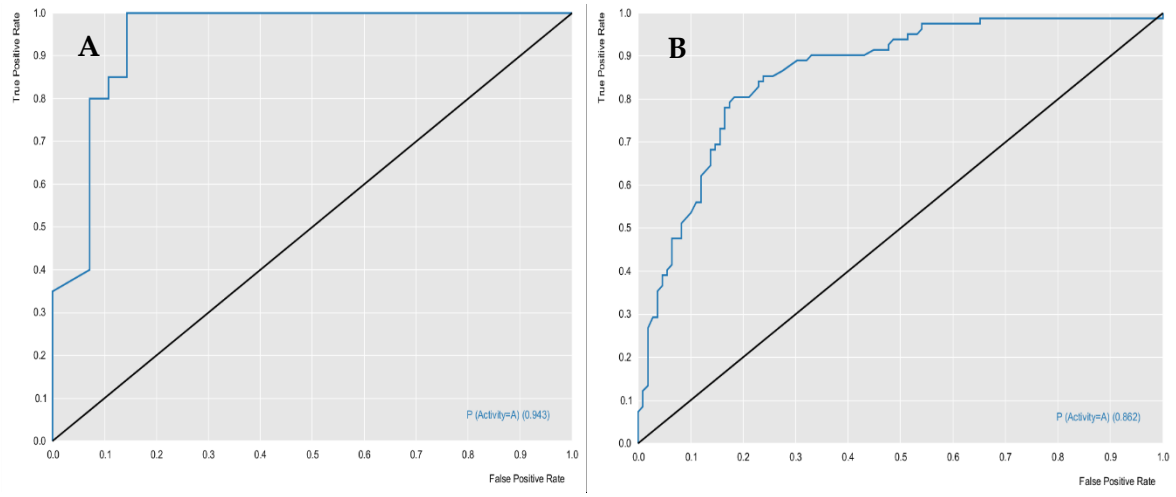


Figure 2. ROC curve generated for the *L. braziliensis* RF model. (**A**) Test and (**B**) cross-validation.

With the models created and demonstrated to have excellent performance, the lignan bank was then screened to select compounds that are potentially active against *L. major* and *L. braziliensis*.

The RF model was able to select 11 compounds with active potential, with probabilities ranging from 50% to 57%, for *L. major* (Table 2), whereas the model for *L. braziliensis* was able to select 21 potentially active compounds, with probabilities ranging between 50% and 75% (Table 3). Compounds 86 and 160 were considered to be most active for *L. major*, and compounds 8, 60, 157, and 160 were considered to be the most active for *L. braziliensis*.

Table 2. Lignans activity probabilities (pActivity) against *L. major*, as assessed by the RF model.

ID	IUPAC Name	pActivity
44	(1R,2S)-2-(2,6-Dimethoxy-4-prop-2-enylphenoxy)-1-(4-hydroxy-3,5-dimethoxyphenyl)propane-1,3-diol	0.54
60	(1S,2S)-1-(4-Hydroxy-3-methoxyphenyl)-2-[4-(3-hydroxypropyl)-2-methoxyphenoxy]propane-1,3-diol	0.53
83	4-[(2S,3R)-5-[(E)-3-Hydroxy-3-methoxyprop-1-enyl]-3-(hydroxymethyl)-7-methoxy-2,3-dihydro-1-benzofuran-2-yl]-2-methoxyphenol	0.55
86	(E)-3-[(2R,3S)-2-(3,4-Hihydroxyphenyl)-3-(hydroxymethyl)-2,3-dihydro-1,4-benzodioxin-6-yl]prop-2-enoic acid	0.57
87	4-[(2R,3R)-3-(Hydroxymethyl)-6-(3-hydroxypropyl)-2,3-dihydro-1,4-benzodioxin-2-yl]-2-methoxyphenol	0.51
124	Ethyl (2S,3R)-2-(3,4-dihydroxyphenyl)-5-[(E)-3-ethoxy-3-oxoprop-1-enyl]-7-hydroxy-2,3-dihydro-1-benzofuran-3-carboxylate	0.52
132	(3R,4S)-4-[(R)-(4-Hydroxy-3-methoxyphenyl)[(5R)-4-hydroxy-5-methoxycyclohexa-1,3-dien-1-yl]methyl]-3-(hydroxymethyl)oxolan-2-one	0.53
157	(2S,3S,4R,5R,6S)-2-{4-[(2S,3R,4S)-4-[(4-Hydroxy-3-methoxyphenyl)methyl]-3-(hydroxymethyl)oxolan-2-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.53
158	(2R,3S,4R,5R,6S)-2-{4-[(1S,3aR,4S,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.50
159	(2R,3S,4R,5R,6S)-2-{4-[(1S,3aR,4R,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.50
160	(2S,3R,4R,5S,6R)-6-{4-[(1S,3aR,4S,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-5-{[(2S,3R,4R)-3,4-dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxy}-2-(hydroxymethyl)oxane-3,4-diol	0.56

Table 3. Lignans activity probabilities (pActivity) against *L. braziliensis*, as assessed by the RF model.

ID	IUPAC Name	pActivity
8	(1S,5S,6R,7S,8R)-7-(1,3-Benzodioxol-5-yl)-1,3,8-trihydroxy-6-methyl-5-prop-2-enylbicyclo[3.2.1]oct-3-en-2-one	0.70
20	4-[[(3R,4R,5S)-5-(4-Hydroxy-3-methoxyphenyl)-4-(hydroxymethyl)oxolan-3-yl]methyl]-2-methoxyphenol	0.60
31	(2R,3R,4R)-3,4-bis[(4-Hydroxy-3-methoxyphenyl)methyl]axonal-2-ol	0.53
32	(2S,3R,4R)-3,4-bis[(4-Hydroxy-3-methoxyphenyl)methyl]axonal-2-ol	0.54
44	(1R,2S)-2-(2,6-Dimethoxy-4-prop-2-enylphenoxy)-1-(4-hydroxy-3,5-dimethoxyphenyl)propane-1,3-diol	0.59
57	(1R,2R)-1-(4-Hydroxy-3-methoxyphenyl)-2-[4-(3-hydroxypropyl)-2-methoxyphenoxy]propane-1,3-diol	0.67
58	4-[3-Hydroxy-2-[1-(4-hydroxy-3-methoxyphenyl)-1-methoxybutan-2-yl]oxypropyl]-2,6-dimethoxyphenol	0.68
59	4-[2-[1-Hydroxy-3-(4-hydroxy-3-methoxyphenyl)propan-2-yl]oxy-1-methoxybutyl]-2,6-dimethoxyphenol	0.64
60	(1S,2S)-1-(4-Hydroxy-3-methoxyphenyl)-2-[4-(3-hydroxypropyl)-2-methoxyphenoxy]propane-1,3-diol	0.70
61	4-[(3S,3aR,6S,6aR)-6-(3,4,5-Trimethoxyphenyl)-1,3,3a,4,6,6a-hexahydrofuro[3,4-c]furan-3-yl]-2,6-dimethoxyphenol	0.64
76	(9R,10R,11R)-11-Hydroxy-3,4,5,19-tetramethoxy-9,10-dimethyl-15,17-	0.52

	dioxatetracyclononadeca-1(19),2(7),3,5,12,14(18)-hexene-6-carboxylic acid	
83	4-[(2S,3R)-5-[(E)-3-Hydroxy-3-methoxyprop-1-enyl]-3-(hydroxymethyl)-7-methoxy-2,3-dihydro-1-benzofuran-2-yl]-2-methoxyphenol	0.62
87	4-[(2R,3R)-3-(hydroxymethyl)-6-(3-hydroxypropyl)-2,3-dihydro-1,4-benzodioxin-2-yl]-2-methoxyphenol	0.5
91	4-[[(2R,3R,4R)-4-[(S)-Hydroxy-(4-hydroxy-3-methoxyphenyl)methyl]-2-methoxyoxolan-3-yl]methyl]-2-methoxyphenol	0.65
121	(2R,3R,4R)-2-(4-Hydroxy-3-methoxyphenyl)-4-[(4-hydroxy-3-methoxyphenyl)methyl]-3-(hydroxymethyl)oxolan-3-ol	0.61
124	Ethyl (2S,3R)-2-(3,4-dihydroxyphenyl)-5-[(E)-3-ethoxy-3-oxoprop-1-enyl]-7-hydroxy-2,3-dihydro-1-benzofuran-3-carboxylate	0.59
156	(2R,3R)-2,3-bis[(4-Hydroxy-3-methoxyphenyl)methyl]butane-1,4-diol	0.57
157	(2S,3S,4R,5R,6S)-2-{4-[(2S,3R,4S)-4-[(4-Hydroxy-3-methoxyphenyl)methyl]-3-(hydroxymethyl)oxolan-2-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.75
158	(2R,3S,4R,5R,6S)-2-{4-[(1S,3aR,4S,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.66
159	(2R,3S,4R,5R,6S)-2-{4-[(1S,3aR,4R,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-6-(hydroxymethyl)oxane-3,4,5-triol	0.66
160	(2S,3R,4R,5S,6R)-6-{4-[(1S,3aR,4S,6aR)-4-(4-hydroxy-3-methoxyphenyl)-hexahydrofuro[3,4-c]furan-1-yl]-2-methoxyphenoxy}-5-{[(2S,3R,4R)-3,4-dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxy}-2-(hydroxymethyl)oxane-3,4-diol	0.72

2.3 Alignment of Protein Sequences

Sequence alignment was used to verify the similarities and identities of the enzymes selected in this study, across different species, which allowed the analysis of conserved regions and the identification of common residues in the active site. In addition, differences and structural similarities could be identified that might contribute to rational drug planning. Therefore, we investigated the shared amino acids in the active sites of various *L. major* and *L. braziliensis* enzymes.

The results showed that most of the enzymes shared greater than 80% identity between *L. major* and *L. braziliensis*, with 84% identity for Glycerol-3-phosphate dehydrogenase (GPDH); 85% for dihydroorotate dehydrogenase (DHODH); 72.13% for the Pteridine reductase 1 (PTR1); 82.69% for Trypanothione reductase (TR); and 85.69% for UDP-glucose pyrophosphorylase (UGPase) (Annex I: Figures S1–S5).

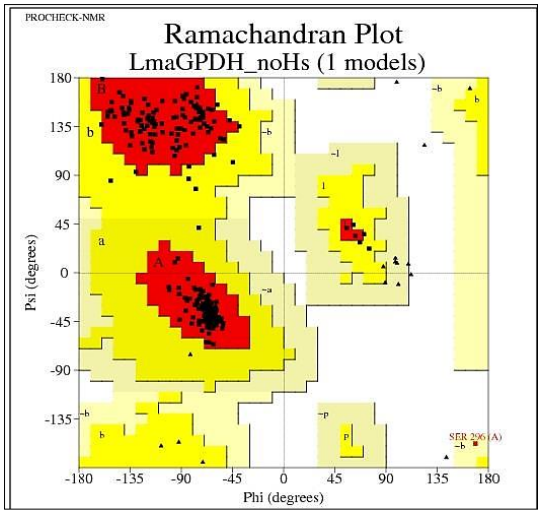
All of the amino acids in the active site of GPDH protein were conserved between species. According to Choe et al. [26], interactions with the residues Trp44, Ile93, Phe101, Phe97, and Met46 of *L. major* GPDH (*Lm*GPDH) were observed for the compound 2-bromo-6-chloro-purine. For the DHODH protein, researchers have described advances and perspectives in structural biology, which were used to identify and validate target sites for the development of drugs that target this enzyme [27]. The study showed that DHODH is

structurally conserved among *Leishmania* species, except in *L. braziliensis*, which showed changes in the amino acid sequence of the active site. In *L. braziliensis*, Met104 is replaced with Ala138, and Cys150 is replaced with Tyr176. In the PTR1 enzyme, interactions were observed between the amino acids Arg17, Ser111, Asp181, Leu188, Leu226, Leu229, His241, Tyr283, and Arg287 and the compound (2~{*R*})-2-[3,4-bis (oxidanyl) phenyl]-6-oxidanyl-2,3-dihydrochromen-4-one, indicating the location of the active site, according to a study performed by Pisa et al. [28]. In this study, we noticed the replacement of the Leu229 residue with Phe229 in the *L. braziliensis* PTR1. The enzymes TR and UGPase showed that the residues in the active site were highly conserved in both species.

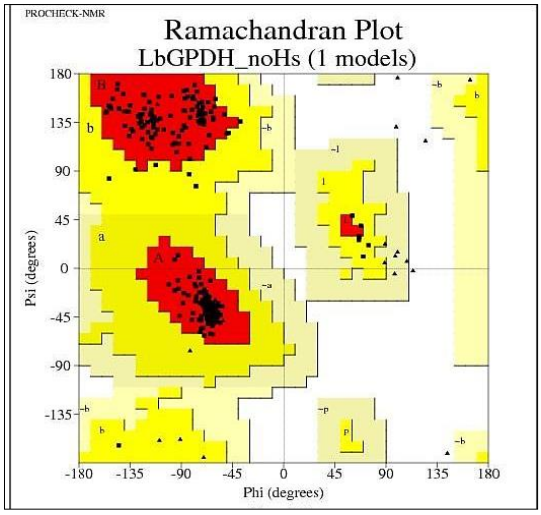
2.4 Homology Modeling

In this study, six models were generated: *L. major* GPDH (*Lm*GPDH) *L. braziliensis* GPDH (*Lb*GDPH), *L. braziliensis* PTR1 (*Lb*PTR1), *L. major* TR *(Lm*TR), *L. braziliensis* TR (*Lb*TR), and *L. braziliensis* UGPase *(Lb*UGPase).

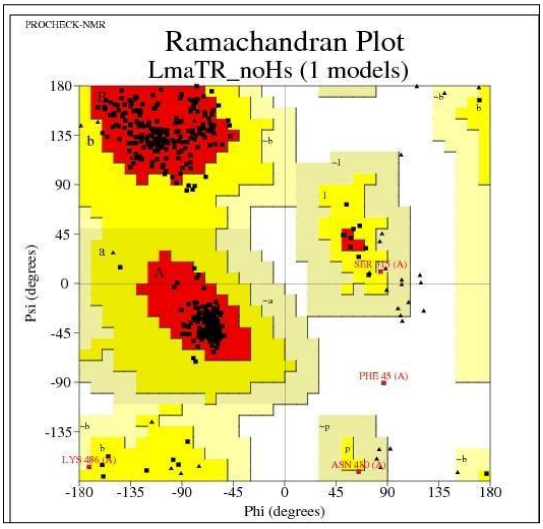
The reliabilities of the models were verified using several tools. One of the primary tools used was the Ramachandran graph, which represents all possible combinations of dihedral angles Ψ (psi) versus φ (phi) for each amino acid of a protein, except for glycine, which has no side chains, and models are considered to be reliable when more than 90% of amino acids are present in the permitted and/or favored regions (colored regions of the graph). Blank regions represent outliers, which have bad contacts. All generated models showed greater than 98% of amino acids in the permitted and favored regions (Figure 3 and Table 4). Highly reliable models were likely obtained due to the high similarity between the model sequences and the high resolution of the templates. All models were used for the following methodologies.



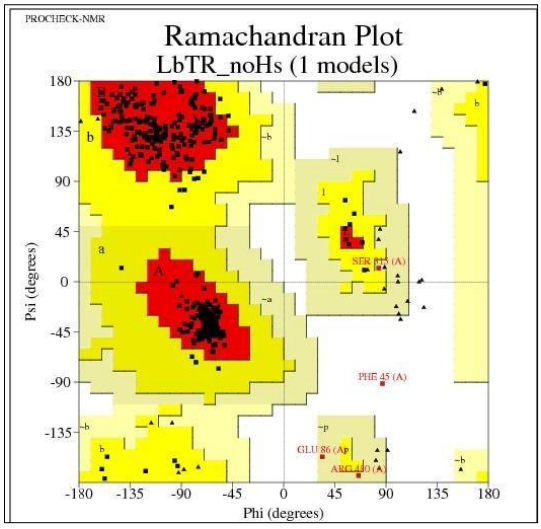
(**A**)



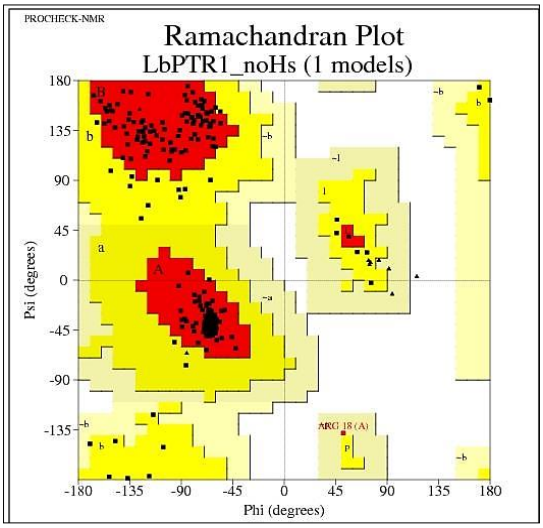
(**B**)



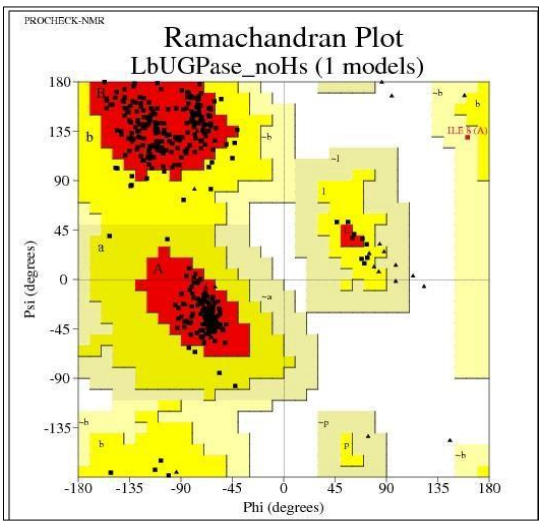
(**C**)



(**D**)



(**E**)



(**F**)

Figure 3. Ramachandran graphs of the homology models generated for *L. major* and *L. braziliensis* enzyme. The colored regions represent the permitted and favored regions of the secondary structures, and the white regions represent the prohibited regions. (**A**) Glycerol-3-phosphate dehydrogenase (GPDH) in *L. major*. (**B**) GPDH in *L. braziliensis*. (**C**) Trypanothione reductase (TR) in *L. major*. (**D**) TR in *L. braziliensis*. (**E**) Pteridine

reductase 1 (PTR1) in *L. braziliensis*. (**F**) UDP-glucose pyrophosphorylase (UGPase) in *L. braziliensis*.

Table 4. Percentage of amino acids present in the permitted and favored regions of the Ramachandran chart for each model.

Enzyme	Species	Ramachandran Percentage
GPDH	*L. braziliensis*	100%
	L. major	100%
TR	*L. braziliensis*	98%
	L. major	98%
PTR1	*L. braziliensis*	100%
UGPase	*L. braziliensis*	100%

Verify 3D analyzes the compatibility of the 3-dimensional (3D) structure with its 1-dimensional (1D) amino acid sequence, based on the characteristics of the chemical environment, such as polarity and compares the results with good structures.

The method determines the environmental characteristics of each residue: (i) The total area of the side chain; ii) the fraction of the side chain area covered by polar atoms or water; and (iii) the local secondary structure. Then, the information is categorized and the structure of the 3D protein is converted into 1D, as a sequence, which represents the environment class of each residue in the structure of the folded protein. Then, the sequence is aligned and compared with sequences of good structures. A reliable model must have a 3D-1D score of more than 80%. All models obtained had scores above 80%, as shown in Table 5.

Table 5. Percentage of the degree of compatibility between the 3D structure and the 1D amino acid sequence, based on Verify 3D generated models.

Enzyme	Species	Verify 3D Percentage
GPDH	*L. braziliensis*	91.83%
	L. major	85.56%
TR	*L. braziliensis*	90.22%
	L. major	93.69%
PTR1	*L. braziliensis*	80.56%
UGPase	*L. braziliensis*	93.47%

The quality of the atomic contacts between the atoms of each residue was analyzed, using the module Coarse Packing Quality Control or Fine Packing Quality Control on the WHAT IF server, which compares the distribution of atom positions around each residue. A

mean score of less than –5.0 indicates bad or unusual atomic contacts. All models generated by the homology presented average score values above −5.0, as shown in Table 6.

Table 6. Average scores for each residue, obtained from the WHAT IF server, for each model.

Enzyme	Species	Average WHAT IF Score
GPDH	*L. braziliensis*	−1.622
	L. major	−1.578
TR	*L. braziliensis*	−0.919
	L. major	−0.894
PTR1	*L. braziliensis*	−0.952
UGPase	*L. braziliensis*	−0.609

2.5 Combined Analysis, Based on Ligand and Structure

The lignans and neolignans that were considered to be active based on the RF models for *L. major* and *L. braziliensis* were submitted to docking consensus evaluation, to increase the reliability of the method and decrease the number of false positives.

In total, 10 enzymes were used, four obtained from the PDB database and six based on homology. The docking results were generated using five different scoring functions and were validated by redocking the PDB ligand with the five types of enzymes for each species. More negative values indicated better predictions for most scoring function, except for the Goldscore and ChemPLP, which rank the best poses using the most positive values.

After docking, the results were standardized, and an average of all energy values was calculated for each lignan. Then, the lignans that obtained lower energy values than the PDB ligand in at least three of the scoring functions were used.

Further calculations were performed to obtain the lignans with the best active potential probabilities for each of the analyzed proteins, using the following formula:

$$Prob = \frac{EM_{Lig}}{EM_{MLig}}, if\ EM_{Lig} < EM_{Inib}$$

where EM_{Lig} is the average energy across all five scoring functions for each analyzed lignan, EM_{MLig} is the highest average energy obtained by the tested lignans, and EM_{Inib} is the average energy across all five scoring functions for the ligand inhibitor, obtained from the crystallographic data of the test protein. Thus, only those compounds that obtained energy values equal to or greater than the interaction energy of the crystallographic inhibitor ligand were considered to be potentially active.

Among the 11 lignans analyzed for *L. major*, all were considered potentially active against GPDH, 10 against DHODH, 9 against PTR1, 9 against TR, and all 11 against UGPase (Annex I: Table S3).

Among the 21 lignans analyzed using the consensus docking analysis for *L. braziliensis*, 20 were potentially active against GPDH, all against DHODH, 19 against PTR1, 16 against TR, and 15 against UGPase (Annex I: Table S4).

Tables S3 and S4 show that the consensus docking analysis returned activity probabilities for the examined lignans that were much higher than that for PDB ligand for various enzymes in the two species.

A second consensus analysis was performed to select multitarget lignans, which demonstrate an active potential probability for more than one protein, in the RF analysis and. For this analysis, the following formula was used:

$$Prob_{Comb} = \frac{(Prob_{Dc} + (1 + ESP) \times P_{Activity}}{2 + ESP}, if\ Prob_{Comb} > 0.5$$

where $Prob_{Dc}$ is the average active potential probability based on the molecular coupling analysis, *ESP* is the specificity value from the RF model, and $P_{Activity}$ is the active potential probability value from the RF model. The combined probability is conditioned, such that only compounds with values above 0.5 are considered likely to be active. The combined probability values were calculated for the lignans for each enzyme studied. Finally, we analyzed which compounds were multitarget compounds. Tables S3 and S4 show the combined probability values between the forecasting models and the molecular docking analysis.

Among the lignin and neolignans bank analyzed in this study, after the combined ligand- and structure-based analysis, and the identification of multitarget compounds, all of the lignans, except lignan 76, were considered to be potentially active against three or more enzymes (Tables 7 and 8). Therefore, 20 molecules were considered to have potential multitarget activity against *L. major* and/or L. *braziliensis*, with 10 molecules common to both species. We also observed that the most likely targets, according to the combined analysis based on structure and ligand, are GPDH, PTR1, TR, and UGPase for *L. major*. While for *L. braziliensis*, the most likely targets are GPDH, PTR1, and UGPase. Figure 4 shows the common lignans for *L. major* and *L. braziliensis* that were considered to be potentially active, based on the RF model, selected by the consensus analysis combined with fit values, and identified as multitarget. We also observed that *L. braziliensis* showed higher

rates of combined probability, indicating that lignans may be potential leishmanicides for this species.

Table 7. Combined probabilities between prediction models and molecular docking analysis for potential activity against *L. major*.

ID	$P_{Activity}$	$Prob_{Comb}$				
		GPDH	**DHODH**	**PTR1**	**TR**	**UGPase**
44	0.54	0.64	0.60	0.65	0.68	0.68
60	0.53	0.66	0.58	0.63	0.66	0.66
83	0.55	0.67	0.59	0.67	0.68	0.68
86	0.57	0.68	0.62	0.68	0.69	0.69
87	0.51	0.65	0.58	-	0.65	0.65
124	0.52	0.56	0.57	0.66	0.69	0.69
132	0.53	0.66	0.57	-	0.66	0.66
157	0.53	0.68	0.63	0.65	-	0.60
158	0.50	0.66	0.59	0.66	-	0.58
159	0.50	0.64	0.63	0.69	0.60	0.66
160	0.56	0.71	0.65	0.71	0.63	0.63

Table 8. Combined probabilities between prediction models and molecular docking analysis for potential activity against *L. braziliensis*.

ID	$P_{Activity}$	$Prob_{Comb}$				
		GPDH	**DHODH**	**PTR1**	**TR**	**UGPase**
8	0.70	0.75	0.66	0.73	-	-
20	0.60	0.69	0.62	0.68	0.66	-
31	0.53	0.65	0.57	0.65	0.62	0.62
32	0.54	0.65	0.57	0.65	0.61	0.62
44	0.59	0.62	0.61	0.71	0.65	0.66
57	0.67	0.72	0.65	0.75	0.69	0.71
58	0.68	0.70	0.65	0.71	0.69	0.72
59	0.64	0.67	0.63	0.69	-	0.70
60	0.70	0.74	0.67	0.75	0.71	0.74
61	0.64	0.71	0.66	0.65	0.68	-
76	0.52	-	0.55	-	-	-
83	0.62	0.72	0.67	0.72	0.70	0.69
87	0.5	0.63	0.55	0.60	0.60	0.59
91	0.65	0.74	0.65	0.73	0.69	0.69
121	0.61	0.71	0.62	0.69	-	-
124	0.59	0.67	0.62	0.70	0.69	-
156	0.57	0.64	0.61	0.66	0.62	0.66
157	0.75	0.79	0.78	0.84	0.77	0.78
158	0.66	0.77	0.68	0.73	0.72	0.72
159	0.66	0.77	0.68	0.75	0.77	0.77
160	0.72	0.76	0.81	0.77	-	0.80

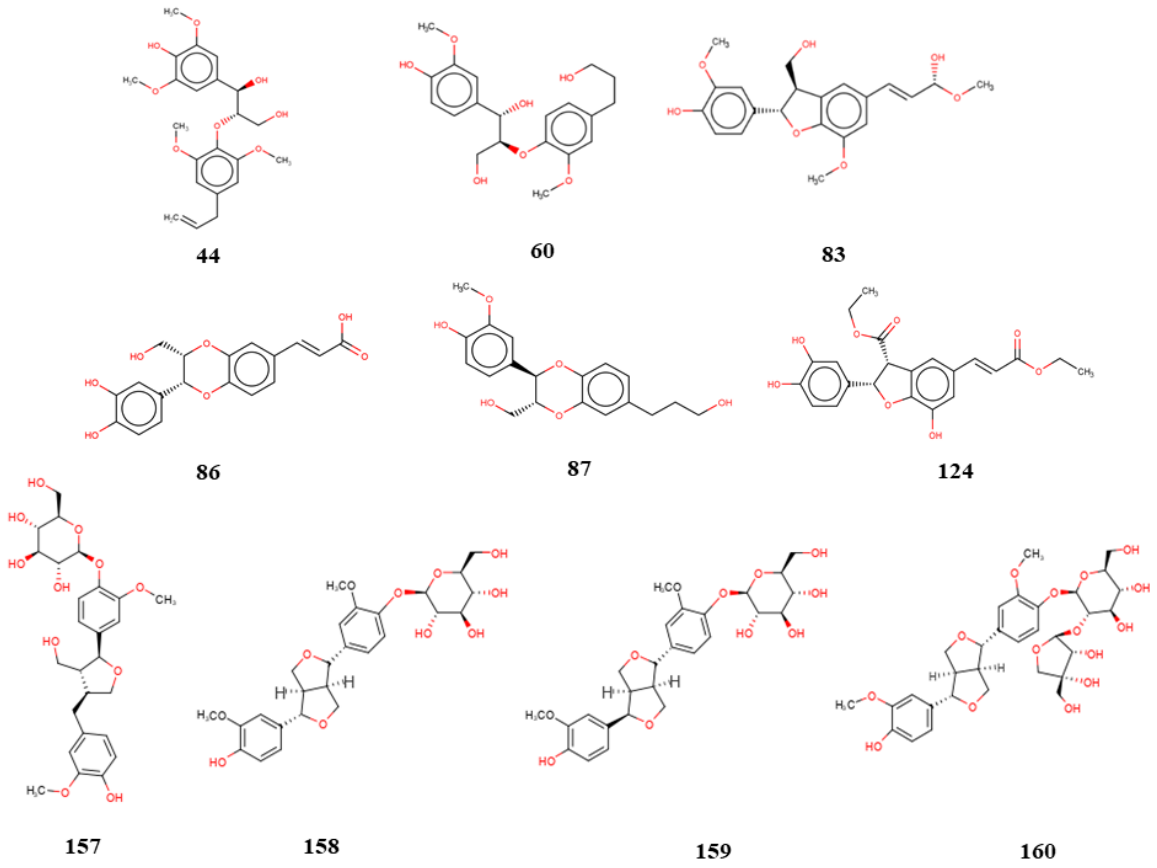


Figure 4. Common compounds that are considered to be potentially active against *L. major* and *L. braziliensis*, based on the random forest model, selected by the consensus analysis, combined with the fit values, and identified as multitarget.

2.5.1 Interaction Analysis

To analyze the interactions, we selected lignan **160** as the compound with the highest leishmanicidal potential because it is considered to be multitarget and obtained activity predictions for Leishmania species (Figures 5 and 6).

GPDH

For the *L. major* GPDH enzyme, the interaction with lignan **160** consisted primarily of large numbers of hydrophobic bonds. Four hydrogen bonds were established, which consisted of bonds with the Ser24, Cys124, Lys126, and Glu301 residues. Four hydrophobic interactions were counted, corresponding with the Thr45, Ile47, Ile94, and Pro95 residues. We observed that although the lignan was located at the enzyme active site, it interacted with other amino acids that are not reported in the literature.

Similar to *L. major*, the *L. braziliensis* GDPH formed four hydrogen bonds with lignan **160**, at the Ser24, Pro95, Thr96, and Glu301 residues. Two hydrophobic interactions, with Trp45 and Ile47, and a steric Val93 interaction were also observed.

DHODH

In *L. major*, lignan 160 formed several hydrogen bonds with the DHODH enzyme active site, providing increased stability to the complex. Links with the amino acids Gly101, Asn107, Ser130, Cys131, Ser130, Pro138, and Val140 and hydrophobic interactions with the amino acids Met104 and Cys150 were observed. According to the literature, lignan **160** interacted with two important amino acids responsible for catalytic activity: Met104 and Cys150.

In *L. braziliensis*, two hydrogen bonds were formed with residues Asn107 and Tyr150, and three hydrophobic interactions were observed with Ser130, Gln139, and Ala146.

PTR1

Hydrogen bonds with the amino acids Ans109, Asn181, and Gly225 were observed between lignan **160** and the active site of *L. major* PTR1. A hydrophobic interaction with Phe113 and a steric interaction with Ser111 were also observed. Important interactions have been reported in the literature between the amino acids Arg17, Ser111, Asp181, Leu188, Leu226, Leu229, His241, Tyr283, and Arg287 and the compound (2~{*R*})-2-[3,4-bis(oxidanyl) phenyl] -6-oxidanil-2,3-dihydrochromen-4-one, indicating that this may be the location of the active site, as reported in a study by Pisa et al. [25]. In the present study, we noticed the substitution of Leu229 residue with Phe229 in *L. braziliensis* PTR1, and a hydrophobic interaction was observed at Phe229. In addition, three hydrophobic interactions, with the amino acids Leu188, Phe191, and Pro230, were also observed. According to the literature [28], the Leu188 residue is conserved in Leishmania species.

TR

The amino acids in the TR active site formed several bonds with lignan 160. Hydrogen bonds were formed with the residues Val55 and Ala363 and hydrophobic interactions were observed with the residues Cys57, Ile199, and Pro336. Interactions with the amino acids Cys57 and Pro336 are considered important for enzymatic inhibition. In contrast, lignan **160** did not obtain good results in *L. braziliensis*.

UGPase

In *L. major*, lignan **160** established hydrogen bonds with the amino acids Arg248, His410, and Pro411. In addition, hydrophobic interactions with residues Arg373 and Val413 and steric interactions with the amino acids Asp62 and Val370 were observed. Most of these amino acids, such as Arg248, Arg373, His410, and Pro411, have been reported in the literature as important for protein inhibition.

In *L. braziliensis*, seven hydrogen bonds were observed between lignan 160 and the UGPase active site, providing increased stability. Links with the amino acids Glu60, Glu62, Ans301, Ser304, Arg374, Ser375, and Leu404 were observed, in addition to a hydrophobic interaction with His411.

2.5.2 Docking Validation

The docking results generated by the five scoring functions were validated by redocking the PDB ligand with the five proteins from *Leishmania* species.

The root-mean-square deviations (RMSDs) of the obtained fitting poses were calculated in comparison with the crystal structure. RMSD values of less than 2 Å indicate an ideal degree of screening reliability. Information regarding the starting structures and the redocking validation results are shown in Table 9.

During the redocking analysis, most of the RMSD values were below 2.0 Å, and all five tested scoring functions positioned the ligand correctly at the active site. The Vina program generated only one ideal RMSD value. Thus, studies that use only the Vina program may generate many false-positive results. In addition, RMSD values for the poses obtained by AD4 could not be calculated because the program does not generate outputs for all ligand poses. Overall, the programs provided values that were considered to be satisfactory for the docking consensus validation.

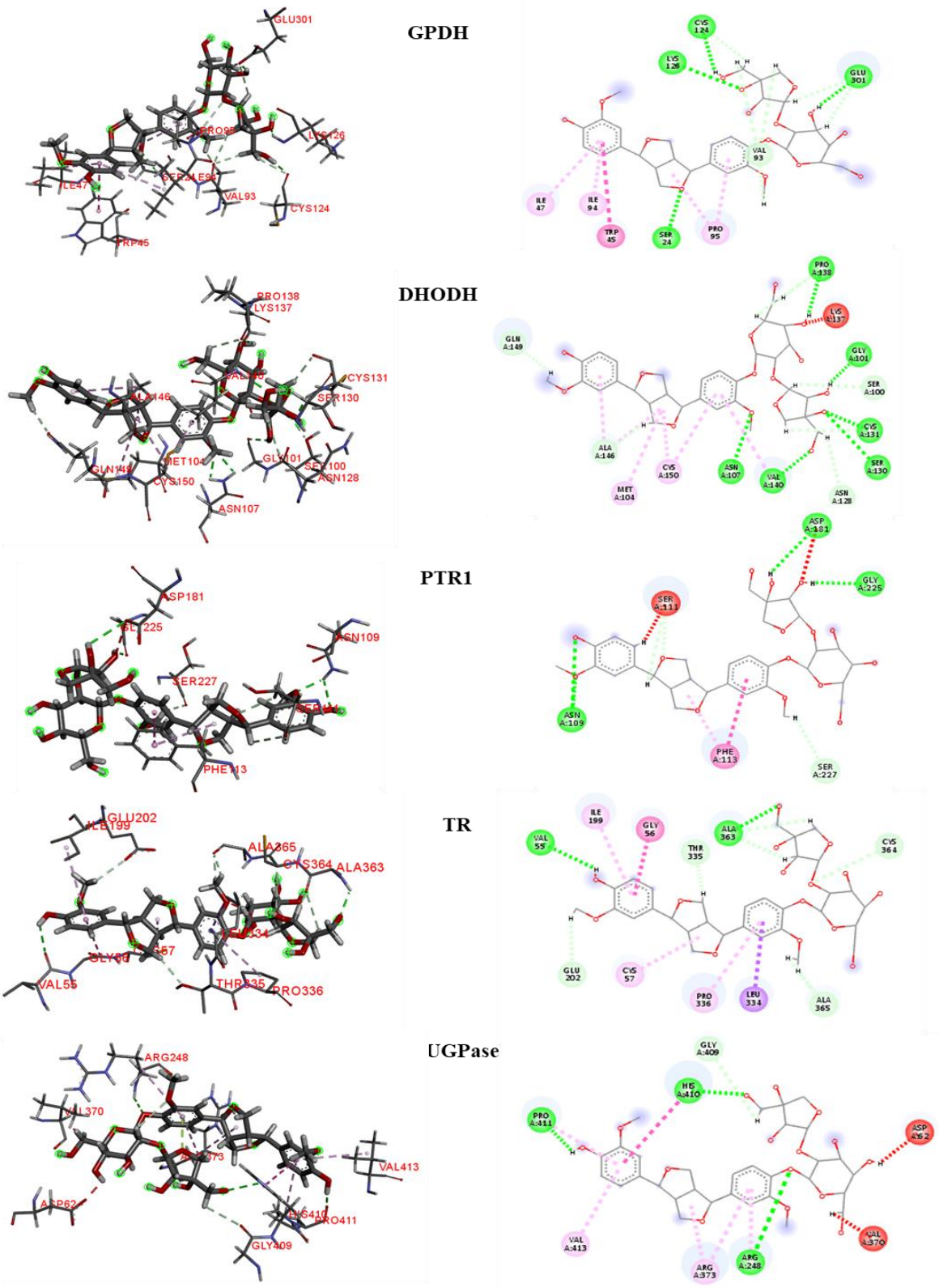


Figure 5. 3D and 2D interactions between lignan 160 and the five examined enzymes [Glycerol-3-phosphate Dehydrogenase (GPDH), Dihydroorotate dehydrogenase (DHODH), Pteridine reductase 1 (PTR1), Trypanothione reductase (TR), and UDP—glucose pyrophosphorylase (UGPase)] in *L. major*. Hydrogen bonds are highlighted in green; hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

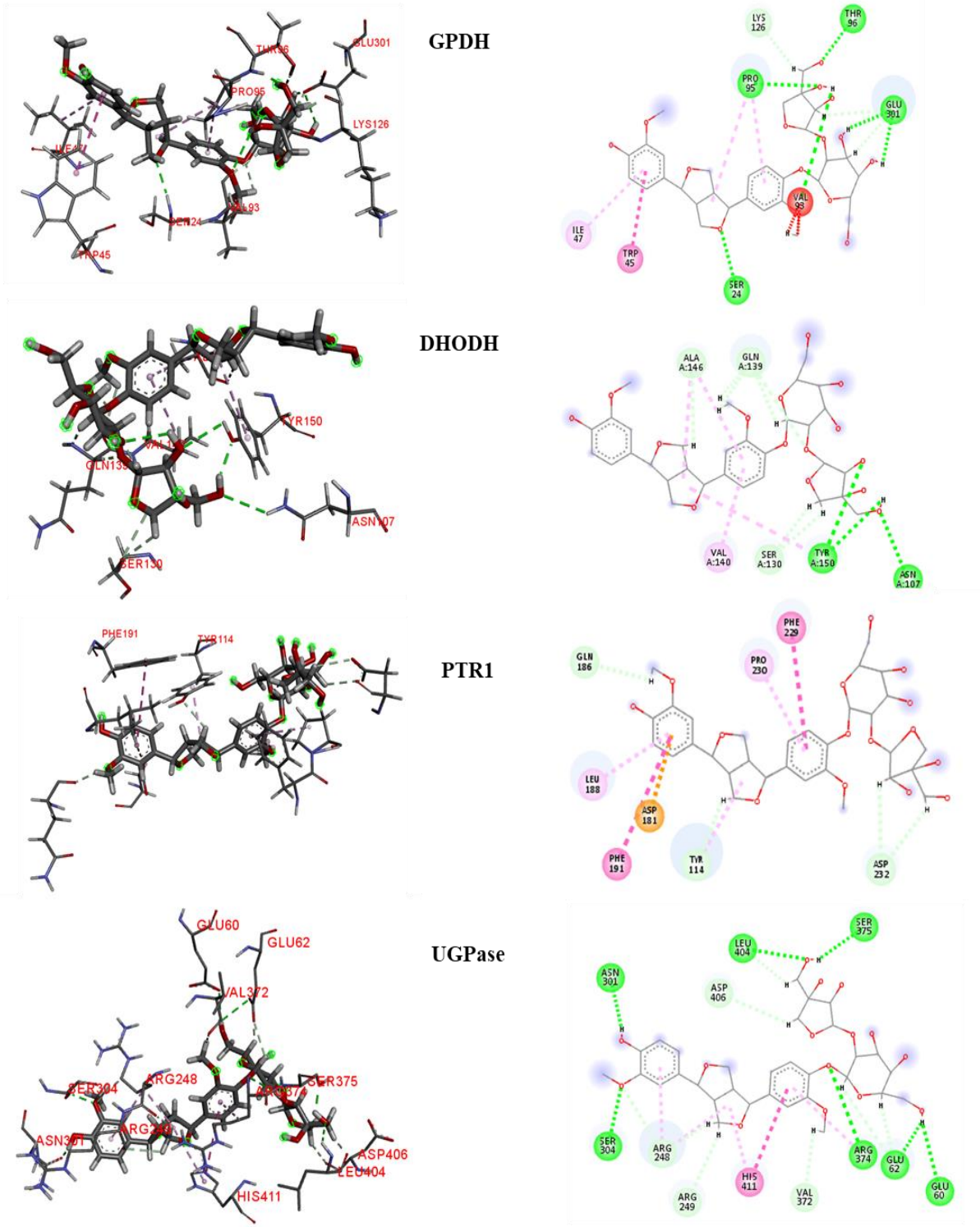


Figure 6. 3D and 2D interactions between lignan 160 and four enzymes [Glycerol-3-phosphate Dehydrogenase (GPDH), Dihydroorotate dehydrogenase (DHODH), Pteridine reductase 1 (PTR1), and UDP—Glucose pyrophosphorylase (UGPase)] in *L. braziliensis*. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

Table 9. Information on the crystalline structures and the root-mean-square deviation (RMSD) values for the poses obtained by redocking.

Protein-Ligand Complex			RMSD				
Enzyme	**PDB ID**	**Inhiibidor**	**Moldoscore**	**Goldscore**	**Vina**	**AD4**	**ChemPLP**
GPDH	1M66	BCP	0.07	0.08	1.95	-	0.21
DHODH	4EF9	4NF	0.23	0.07	3.77	-	0.05
PTR1	5L42	6J6	0.05	0.19	2.50	-	0.36
TR	5EKB	RDS	0.20	0.02	9.66	-	0.49
UGPase	5NZM	9ET	0.29	0.15	2.05	-	0.13

2.5.3 Evaluation of Docking Programs

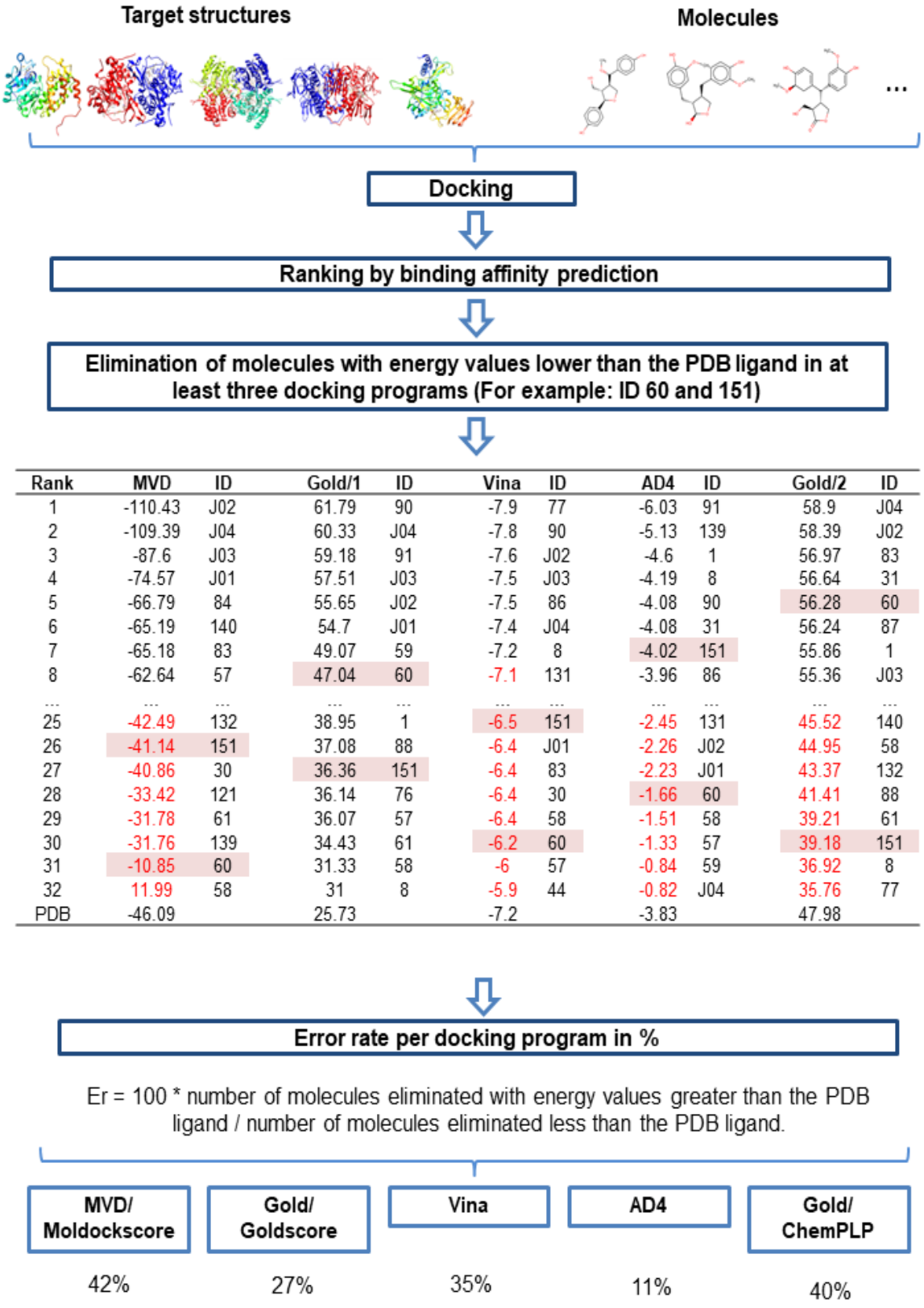
The performances of the programs used for docking analysis in this study were evaluated, by analyzing how each program ranked the compounds that were discarded from the study for having lower energy values than the PDB ligand in at least three docking programs. The workflow used to perform the step-by-step calculations of program error rates (E_r) is shown in Figure 7. The calculation of E_r was used to analyze the probability that a given program would classify an inactive molecule as active, verifying its performance.

The results showed that the MVD, Gold, Vina, AD4, and Plants programs presented E_r values of 42%, 27%, 35%, 11%, and 40%, respectively. Therefore, AD4 had the highest hit rate (89%). Although AD4 is more restrictive for the selection of active compounds, the program was able to obtain excellent results when compared against other programs. Table 10 shows the E_r value *per* program/enzyme, revealing that MVD had a higher total E_r, with higher E_r values identified for DHODH and UGPase compared with those for the other programs.

Table 10. Error rate and hit rate, calculated for each docking program, by target.

Enzyme	Discarded Molecules	Scoring Functions				
		Moldocksocore	**Goldscore**	**Vina**	**AD4**	**ChemPLP**
GPDH	8	0	8	8	1	6
DHODH	48	26	10	25	16	20
PTR1	57	16	18	15	2	14
TR	58	21	10	24	4	35
UGPase	35	24	12	1	1	8
E_r *		42%	27%	35%	11%	40%
H_r *		58%	73%	65%	89%	60%

E_r—error rate; H_r—hit rate.



Rank	MVD	ID	Gold/1	ID	Vina	ID	AD4	ID	Gold/2	ID
1	-110.43	J02	61.79	90	-7.9	77	-6.03	91	58.9	J04
2	-109.39	J04	60.33	J04	-7.8	90	-5.13	139	58.39	J02
3	-87.6	J03	59.18	91	-7.6	J02	-4.6	1	56.97	83
4	-74.57	J01	57.51	J03	-7.5	J03	-4.19	8	56.64	31
5	-66.79	84	55.65	J02	-7.5	86	-4.08	90	56.28	60
6	-65.19	140	54.7	J01	-7.4	J04	-4.08	31	56.24	87
7	-65.18	83	49.07	59	-7.2	8	-4.02	151	55.86	1
8	-62.64	57	47.04	60	-7.1	131	-3.96	86	55.36	J03
...	...	...	...	...	...	...	...	...	...	...
25	-42.49	132	38.95	1	-6.5	151	-2.45	131	45.52	140
26	-41.14	151	37.08	88	-6.4	J01	-2.26	J02	44.95	58
27	-40.86	30	36.36	151	-6.4	83	-2.23	J01	43.37	132
28	-33.42	121	36.14	76	-6.4	30	-1.66	60	41.41	88
29	-31.78	61	36.07	57	-6.4	58	-1.51	58	39.21	61
30	-31.76	139	34.43	61	-6.2	60	-1.33	57	39.18	151
31	-10.85	60	31.33	58	-6	57	-0.84	59	36.92	8
32	11.99	58	31	8	-5.9	44	-0.82	J04	35.76	77
PDB	-46.09		25.73		-7.2		-3.83		47.98	

Figure 7. Workflow used to verify the performance of the docking programs, using the connection energy values.

2.6. Prediction of Drug Resistance

One of the great justifications for the development of new drugs against Leishmaniasis is the resistance of some species to commercialized drugs. To minimize the possible effects of resistance to a likely drug candidate, we searched the TritryDB database for single nucleotide polymorphisms (SNP). SNPs are mutations that are frequently present in over 1% of the species and may contribute to the development of drug resistance.

The results showed that only one SNP was identified in *L. major* PTR1, whereas in *L. braziliensis*, non-synonymous SNPs were identified in both DHODH and TR. No data were found for UGPase in either species.

Among the 14 SNPs identified, only four presented a polymorphic allele with relevant allele frequency, between 40% and 50%, as shown in Table 11.

After identifying the primary SNPs that can cause drug resistance, we located the ancestral amino acids in the enzyme structure that are likely to be replaced by SNPs and determined whether these changes were near or in the active site of the target protein. We found that none of these amino acids were located near the active site of the studied proteins, which reduces the likelihood of drug resistance (Figure 8).

Table 11. List of single nucleotide polymorphisms (SNPs) identified in the TritryDB database, with information regarding the ancestral allele, polymorphic allele, allelic frequency, and amino acid position for each species and enzyme studied. The SNPs with the highest allelic frequencies are highlighted in bold.

Species	ID TritryDB	Non-Synonymous SNP	Allele/Amino Acid	Allele Frequency	Protein Position
DHODH					
L. major	LMJSD75 16001070	-	-	-	-
L. braziliensis	LbrM.16.0550	A T A	G (Ala)/**A (Thr)** T (Val)/C (Ala) G (Gly)/**A (Ser)**	0.50/**0.50** 0.75/0.25 0.50/**0.50**	301 205 6
GPDH					
L. major	LmjF.10.0510	C	G (Arg)/C (Pro)	0.83/0.17	180
L. braziliensis	LbrM.10.0640	-	-	-	-
PTR1					
L. major	LmjF.23.0270	A	G (Gly)/**A (Glu)**	0.60/**0.40**	25
L. braziliensis	LbrM.23.0300	-	-	-	-

TR					
L. major	LMJLV39 050008400	-	-	-	-
L. braziliensis	LbrM.05.0350	T	C (Ala)/T (Val)	0.67/0.33	36
		G	A (Thr)/G (Ala)	0.67/0.33	97
		G	A (Asp)/G (Gly)	0.67/0.33	112
		G	A (Asn)/G (Ser)	0.67/0.33	116
		C	G (Glu)/C (Asp)	0.67/0.33	115
		G	A (Asn)/G (Ser)	0.67/0.33	116
		A	C (Gln)/A (Lys)	0.67/0.33	130
		A	**A (Asn)**/G (Leu)	**0.50**/0.50	144
		C	G (Lys)/C (Asn)	0.67/0.33	480
UGPase					
L. major	LMJLV39 180015400	-	-	-	-
L. braziliensis	LbrM.18.1050	-	-	-	-

2.7. Molecular Dynamics Simulations

We selected some lignans that were considered to have potential multitarget leishmanicidal activity against the studied species, with excellent ADMET properties, and examined them using molecular dynamics (MD) simulations. MD assesses the flexibility of enzymes and the stability of interactions, depending on the conditions, such as solvent used, ion concentration, pressure, and temperature. Therefore, the interactions between lignans and the crystallographic ligands were used to study the flexibility and conformational changes that affect the complexes during the MD simulation. RMSD was calculated for the Cα atoms in each complexed enzyme and the structures of each ligand, separately.

The RMSD analysis of the complexed GPDH enzyme showed conformations ranging from 0.2 to 0.4 nm in size during 10 ns, with high stability (Figure 9A), except for GPDH complexed with compound 83, which showed a peak of instability at 5 ns but was quickly stabilized. The same pattern was observed for DHODH (Figure 9B). PTR1 also showed stability during 10 ns, except for the protein complexed with compound 159 (Figure 9C). The PTR1-157 complex was considered to be the most stable. TR complexed with lignan 20 showed greater stability than the complex containing the crystallographic ligand (Figure 9D). All complexes with the UGPase protein showed stability after 2 ns (Figure 9E). However, it was not possible to perform MD calculations for the crystallographic ligand of DHODH, due to problems with the parameterization of the ligand.

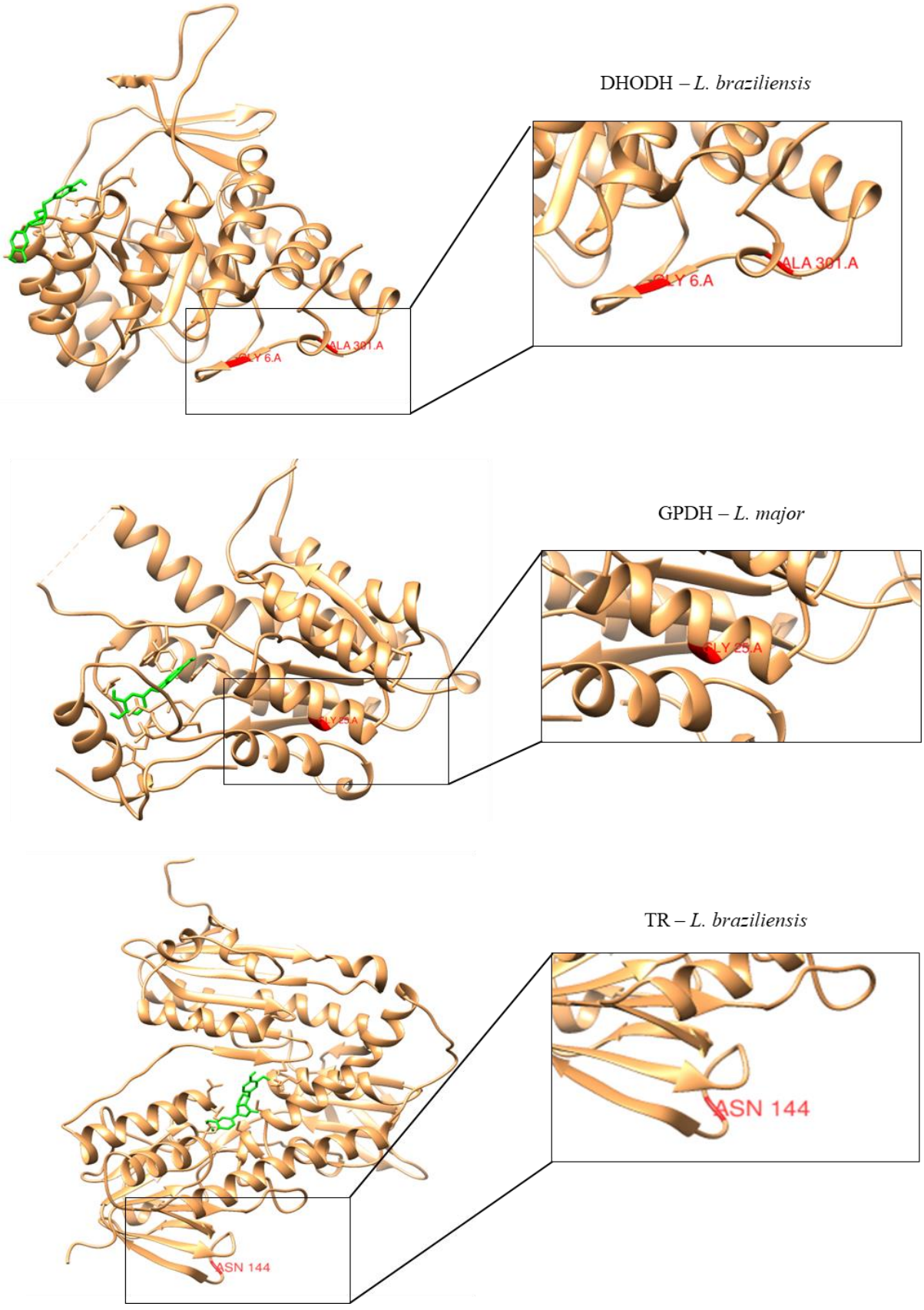
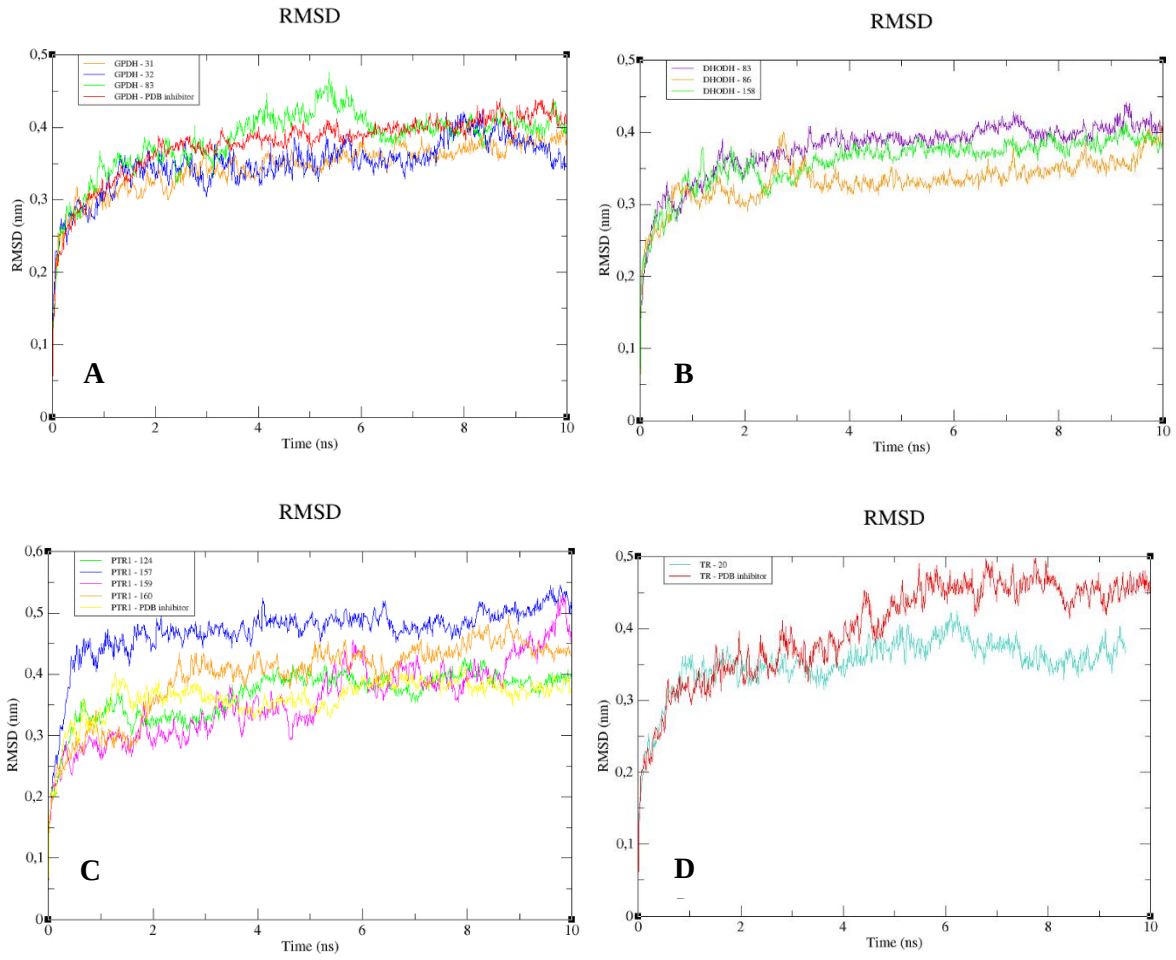


Figure 8. Three-dimensional structure of *L. major* and *L. braziliensis* enzymes. Amino acids that are likely to be affected by SNPs are highlighted in red. Compounds located in the active site of their respective enzymes are highlighted in green.

When we analyzed the flexibility of the ligands, we found that compounds 31, 83, and the PDB ligand complexed with GPDH enzyme were more stable (Figure 10A). The same pattern was also found for complexes with DHODH (Figure 10B). For the PTR1 enzyme, lignan 157 showed high instability, unlike the other molecules, which remained stable (Figure 10C). Small peaks of instability were observed for TR compounds (Figure 10D). Compounds 86 and the PDB ligand showed stability during 10 ns for the UGPase protein (Figure 10E). In addition, lignan 44 achieved stability from 2 ns and lignan 160 achieved stability from 4 ns for UGPase.



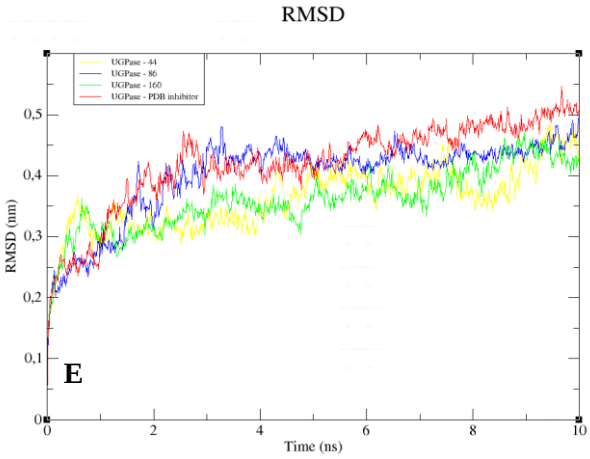
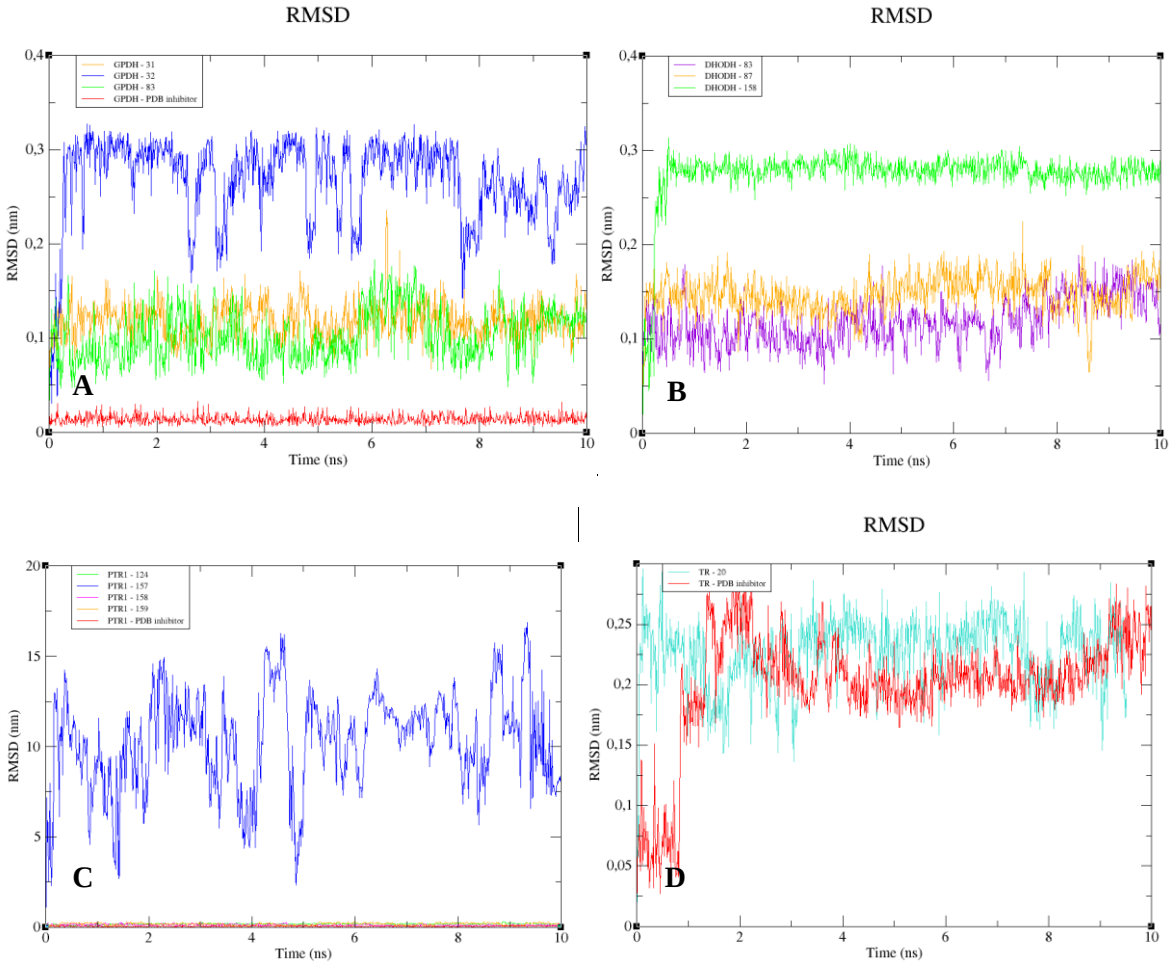


Figure 9. RMSD values for the Cα atoms of enzymes complexed to lignans and the Protein Data Bank (PDB) ligand. (**A**) GPDH. (**B**) DHODH. (**C**) PTR1. (**D**) TR. (**E**) UGPase.



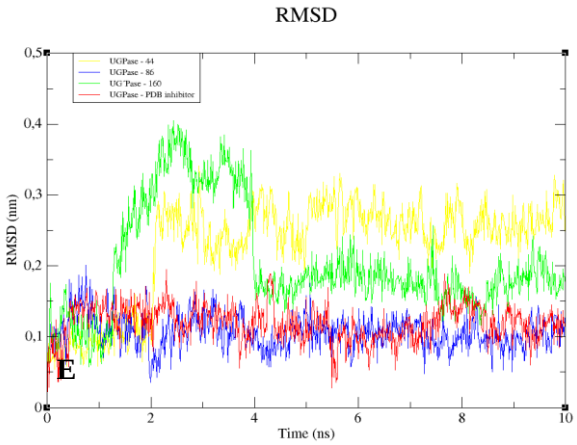


Figure 10. The RMSD values of the Cα atoms of the lignans and the PDB ligand. (**A**) GPDH. (**B**) DHODH. (**C**) PTR1. (**D**) TR. (**E**) UGPase.

To understand the flexibility of the residues and the amino acids that contribute to the conformational changes in the enzymes, the root-mean-square fluctuation (RMSF) values for each amino acid in each enzyme were calculated. High RMSF values suggest increased flexibility, whereas low RMSF values reflect decreased flexibility. Given that amino acids with fluctuations above 0.3 nm contribute to the flexibility of the protein structure, we found that residues at positions 1, 182, 285–290, 338–340, and 345–349 contribute to conformational changes in GPDH, with only one residue from the active site complexed to compound **83**, favoring the alteration (Figure 11A). Among the more than 300 amino acids present in DHODH, only amino acids 1, 10–25, 241–243, 250, 347, 348, and 351–354 contribute to conformational changes (Figure 11B). In PRT1, we observed that residues 1–4, 76–79, 120–123, 125–130, 132–134, 168, and 231 favor structural changes (Figure 11C). In TR, amino acids 74–88, 91, 173, 186, 351, 360, 403, 447, 448, 450, 452, 453, 456, and 458–463 favor structural changes (Figure 11D). In UGPase, the amino acids 6, 7, 43, 53, 171, 180, 201, 265, 267–274, 342, 343–346, 348, 467, 468, 487, and 488 (Figure 11E) favor structural changes. We found that none of the amino acids that affect structural conformations identified in DHODH, PTR1, TR, and UGPase are components of the active site.

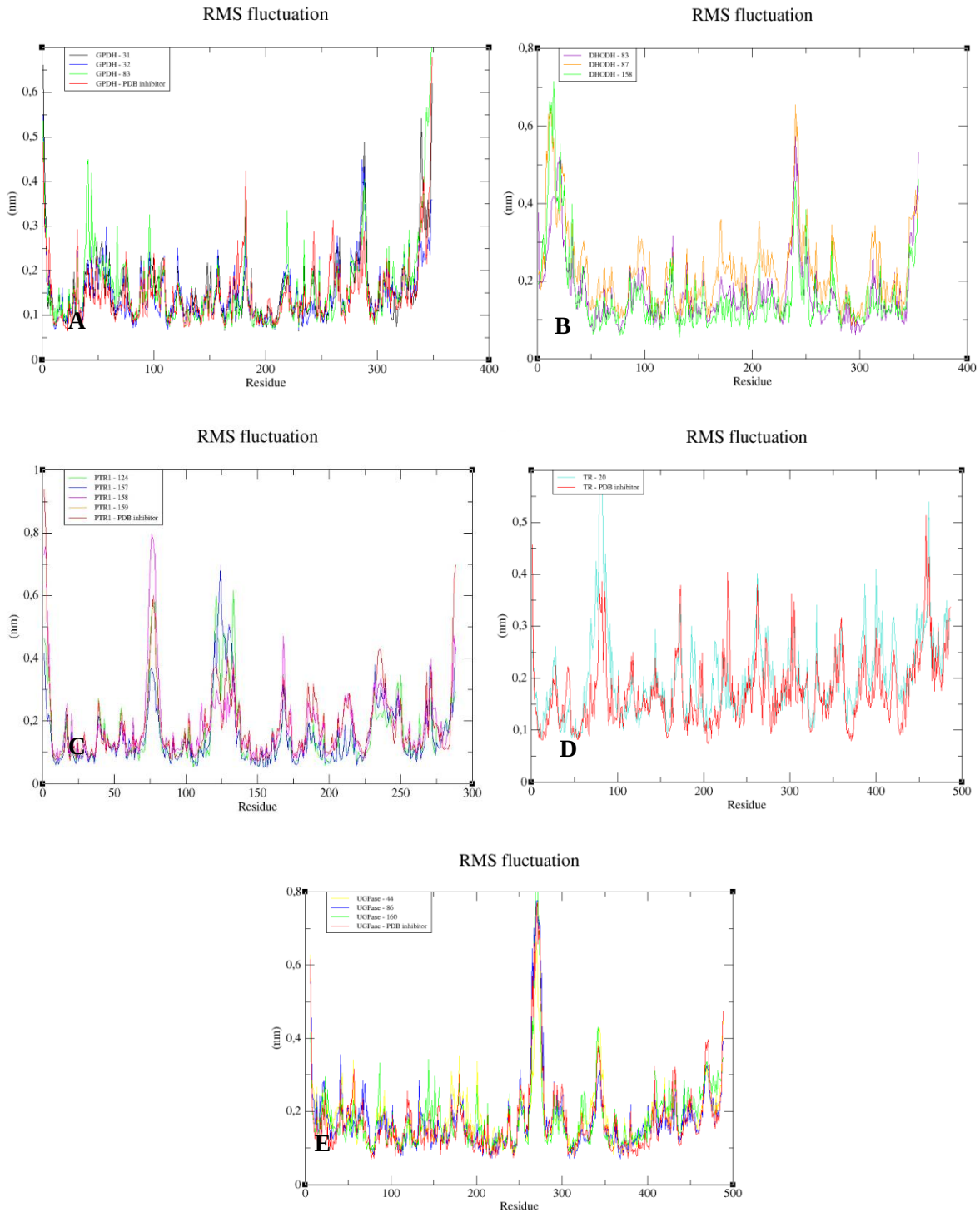


Figure 11. Root-mean-square fluctuation (RMSF) values for the Cα atoms of enzymes complexed with lignans and the PDB ligand. (**A**) GPDH. (B) DHODH. (**C**) PTR1. (**D**) TR. (**E**) UGPase.

2.8 Activity Test against *L. major* and *L. braziliensis* Axenic Promastigotes

We selected four lignans that achieved excellent results during the virtual screening process and isolated them. The compounds were evaluated for their potential to inhibit the growth of promastigotes forms of *L. major* and *L. braziliensis*. Only compound **156** was not tested in *L. major* because it did not obtain good results in silico. The results showed that lignan (**159**), epipinoresinol-4-O-β-D-glucopyranoside, displayed antileishmanial activity against *L. major* promastigotes, with an inhibitory concentration (IC_{50}) of 36.51 µM (Table 12). When investigated *against L. braziliensis* promastigotes, compounds (**156**) secoisolariciresinol, (**158**) pinoresinol-4-O-β-D-glucopyranoside, (**159**) epipinoresinol-4-O-β-D-glucopyranoside, and (**160**) pinoresinol-4-O-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside inhibited growth with IC_{50} values of 9.28, 36.35, 5.39, and 13.77 µM, respectively (**Table 11**). The results showed that compounds **156**, **158**, and **159**, showed excellent potential as growth inhibitors during the promastigote stage. In addition, we observed that compounds (**158**), pinoresinol-4-O-β-D-glucopyranoside and (**159**), epipinoresinol-4-O-β-D-glucopyranoside, which are epimers, showed significantly different biological activities, with compound (**159**) epipinoresinol-4-O-β-D-glucopyranoside having the greatest potential against *L. braziliensis*. This result confirms the docking study consensus and combined analysis, which revealed greater potential activity for molecule **159** than for molecule **158**.

Table 12. Antileishmanial activity of lignans against *L. major* and *L. braziliensis* promastigotes.

ID	Name	IC_{50} (µM)	
		L. major	***L. braziliensis***
156	Secoisolariciresinol	-	9.28
158	Pinoresinol-4-*O*-β-D-glucopyranoside	>50	36.35
159	Epipinoresinol-4-*O*-β-D-glucopyranoside	36.51	5.39
160	Pinoresinol-4-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside	>50	13.77
	Meglumine antimoniate	>40	>40
	Amphotericin B	12.4	18

3. DISCUSSION

The set of 160 Lignans and Neolignans from different subclasses (14 Furans, 10 Furofurans, 14 Dibenzylbutyrolactols, 22 Dibenzylbutanes, 21 Dibenzocycloocyadienes, 17

Aryltetralins, 3 Arylnaphthalenes, 8 Neolignans alkyl aryl ethers, 16 Neolignans benzofurans, and 9 Neolignans benzodiones) were used in this study. Lignans and Neolignans comprise a class of secondary metabolites, with diverse chemical structures, that are found in more than 70 families of plants and exhibit several significant and potent biological activities, including antioxidant [21,22], anti-inflammatory, hepatoprotective, anticancer [22,23], antimicrobial [22], trypanocidal [24,25], neuroprotective [26], and larvicidal [27] activities.

Initially, the set of 160 Lignans and Neolignans were used by various predictive parameters, which were investigated using ADMET profiles. Thus, compounds with better ADMET profiles were subjected to further virtual screening methodologies.

The RF model was able to select 11 compounds with active potentials ranging from 50% to 57%, for *L. major*, and 21 potentially active compounds for *L. braziliensis*, with probabilities ranging between 50% and 75%. We observed the probabilities for biological activity were higher for *L. braziliensis;* therefore, we suggest that lignans may offer greater therapeutic potential against this species.

Virtual ligand-based screening is a method that is capable of evaluating and/or selecting compounds with desired properties, using chemical structures associated with known biological activity data to develop models, such as QSAR analyses [29]. QSAR models contribute to planning and drug development by reducing the costs of new molecule development and reducing the number of animals necessary for experimental tests.

Docking is a virtual screening method, based on structure that can identify selective compounds and predict the mechanisms of action. We performed consensus docking for the molecules that were identified as potentially active by the RF models. The consensus docking method allows the elimination of potential false positives and ensures the increased reliability of the procedure. The results showed that all of the lignans identified as potentially active by the RF models achieved excellent results during the consensus docking analyses for different enzymes.

When analyzing the results of consensus docking, we observed that the same compounds were ranked among the first five positions in one docking program, whereas it was listed among the bottom positions in another program. In addition, when standardizing the docking results, we identified a compound that failed in three programs; however, when all the binding energies were averaged, that compound increased in rank. These results were due to a single program assigning this molecule much higher energy values than the other programs. Therefore, we used two criteria for the elimination of compounds during the consensus docking analysis: i) Compounds that failed in at least three docking programs; and ii)

compounds that obtained average consensus docking energy values below that of the crystallographic ligand.

Understanding the performance of each docking program is necessary to verify the quality and reliability of the virtual screening process, regardless of whether a consensus approach is applied. Therefore, the use of a consensus strategy and the performance evaluation of the docking programs were essential for our structure-based virtual screening analysis.

A study performed by Chang et al. [30] compared the virtual screening results obtained for HIV protease inhibitors, using AutoDock 4 and Vina. The authors concluded that both AutoDock 4 and Vina were able to select active compounds (AUC = 0.69 and 0.68, respectively; $p < 0.001$) and that Vina was more scalable for the treatment of difficult coupling problems (i.e., larger and more flexible compounds) than AD4. Another study, recently performed by Ren et al. [31], investigated the performance of several docking programs, to validate a scoring function strategy, and observed that among the evaluated programs, Plants and Vina obtained success rates of 53.6% and 48.5%, respectively. In contrast, the Gold program achieved success rates of up to 81% (within 2.0 Å of the experimental binding mode), according to Verdonk et al. [32] and was more than 60% successful at predicting the correct poses for 84 evaluated protein-ligand complexes, in a study by Chaput and Mouawad [33].

The combined analysis, based on both ligand and structure, allowed the selection of several potentially active and multitarget compounds against three or more enzymes, except lignan **76**, which obtained good results against only two enzymes. We noticed that the combined probability values were better for the species *L. braziliensis*, for which most of the lignans showed $Prob_{Comb}$ values between 60% and 80% of activity. These results were confirmed by in vitro assays, using *L. braziliensis* promastigotes, as all tested molecules inhibited parasite growth. In *L. major*, of the four compounds tested, only one achieved antileishmanial activity. We, therefore, suggest that a probability of activity greater than 60% is necessary to inhibit growth.

An interesting study by Stevanovic et al. [34] found success in their study using homology detection methods and in silico screening to search for potential inhibitors of a new target, type 2 NADH dehydrogenase of *Leishmania infantum*. According to the researchers, the selected compounds that exhibited favorable properties in the computational screening experiments were tested and the leishmanicidal activity was determined in amastigotes and wild-type axenic promastigotes of *L. infantum*. The results showed that the identified

compound, a substituted 6-methoxy-quinalidine, showed promising activity under the two cellular forms.

Although the isolated compounds have been tested only in amastigote forms, several studies indicate that the enzymes addressed in the study are expressed in amastigote and promastigote forms and there is proven biological activity for these enzymes in both forms. According to Choe et al. [26], the enzyme GPDH is more expressed especially in the amastigote form, where fatty acids, instead of carbohydrates, are predominantly used as an energy source. According to Steiner et al. [35], glucose UDP is essential for all organisms and in *Leishmania*, several gluconjugates are expressed during the parasite's life cycle, allowing survival and proliferation in the vector and in the mammalian host. Experimental data for the in vitro inhibition of *Lm*DHODH indicate that natural products can actually inhibit *Lm*DHODH against promastigotes and amastigotes [36]. The same study identified several secondary metabolites that were able to inhibit *Lm*DHODH in vitro at concentrations of IC_{50} 27 μM, 30 μM and 31 μM. Compounds based on a structure-guided approach designed to have anti-leishmanial activity through the anti-folate mechanism, targeting *Lm*PTR1 in vitro was promising for promastigote and amastigote forms with IC_{50} values of 4.2 μM and 3.3 μM, respectively [37]. Da Sila et al. [38] reported inhibitory activity of the compound ResAn2 for the target TR of *L. braziliensis* against promastigotes and amastigotes with IC_{50} values of 10.27 μM and 17.54 μM.

We also observed that three of the four isolated lignans, which showed inhibitory activity for *L. braziliensis*, belong to the furofurans class. In addition, these lignans have glucose units attached to their structures, which can enhance their therapeutic action. According to Xu et al. [39], furofurans are lignans originally formed by the enantioselective dimerization of two units of coniferyl alcohol derived from the biosynthetic shiquimate pathway. Furofuran lignans are known to have a diversity of structures due to bonding patterns, different substituents, and different configurations [39]. This diversity contributes to a variety of biological activities, including anti-cancer [40], antioxidants, anti-inflammatory, cytotoxic, antimicrobial [39], and antiestrogenic [41]. Therefore, this research brings new information about antileishmanial activity for this class of lignans.

4. MATERIALS AND METHODS

Figure 12 shows a schematic depicting all of the methodologies used in this study.

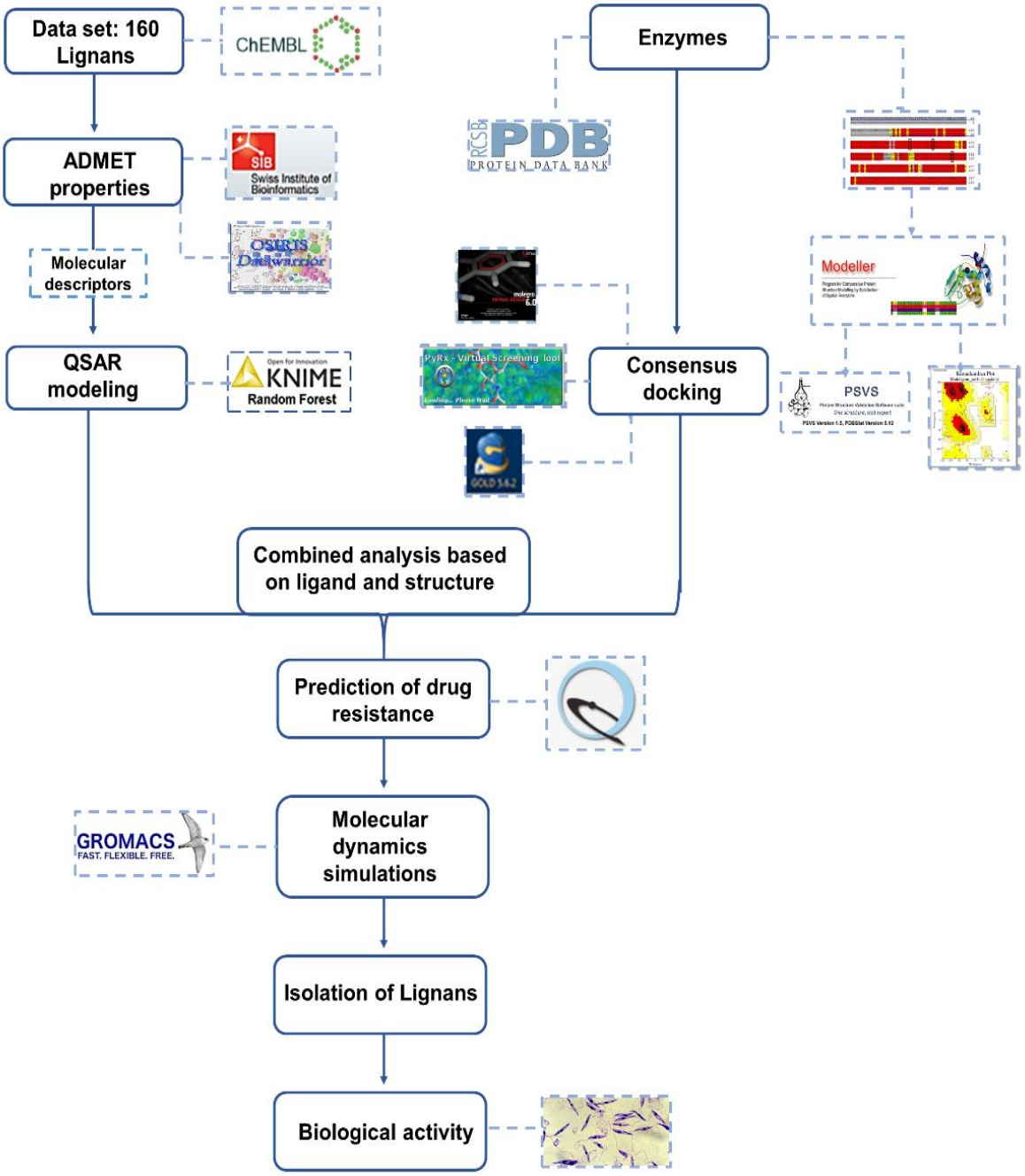


Figure 12. Scheme of all procedures used in this study.

4.1 Predicting ADMET Properties

ADME parameters were calculated using the SwissADME open-access web tool (Swiss Institute of Bioinformatics, Switzerland, http://www.swissadme.ch), whereas the toxicity prediction was performed in the OSIRIS Property Explorer (Idorsia Pharmaceuticals Ltd, Allschwil, Switzerland, https://www.organic-chemistry.org/prog/peo/) [42]. For absorption, factors including membrane permeability and intestinal absorption were considered. We also

investigated compounds that did not exceed more than two violations of the Lipinski rule and for which the logP consensus value was not greater than 4.15. The distribution was assessed by factors that included the blood–brain barrier (logBB) and the permeability of the central nervous system (CNS). Metabolism was predicted based on the substrate models of cytochrome P (CYP). Compounds that were substrates or inhibited more than two enzymes (CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4) were eliminated. The toxicity of the drug was predicted based on the following parameters: Mutagenicity, tumorigenicity, reproductive effects, and irritability.

4.2 Data Collection and Curation

Chemical compounds with known activity (pIC_{50}) against *L. major* (CHEMBL612879) and *L. braziliensis* (CHEMBL612878) (EMBL-EBI, Wellcome Genome Campus, Cambridgeshire, England) were obtained from the ChEMBL database (https://www.ebi.ac.uk/chembl/) [43] for the construction of predictive models. In addition, we obtained 160 lignans e neolignans from ChEMBL to use during virtual screening for the identification of compounds with leishmanicidal potential. All compounds were selected for chemical and biological data, according to the workflows established by Fourches et al. [44,45]. A duplicate search was performed using the HiT QSAR software (Hierarchical QSAR technology, Ukraine). The 3D structures were generated by ChemaxonStandardizer v.18.17.0, (ChemAxon, Boston, USA, www.chemaxon.org).

4.3 QSAR Modeling

Knime 3.6.2 software (Knime 3.6.2, Copyright Miner, from Konstanz Information, Zurich, Switzerland, www.knime.org) was used to perform QSAR modeling. Given the success of our previous studies [46,47], we opted to perform a QSAR 3D analysis. For this, all compounds with a solved chemical structure were saved in SDF format and imported into Dragon 7.0 software (Kode Chemoinformatics srl, Pisa, Italy) [48], to generate descriptors. The RF algorithm was used to build prediction models. The applicability domain was estimated, according to procedures described [49]. External cross-validation was performed, to estimate the predictive power of the models developed. In addition, the performance of external models was assessed by ROC analysis. The models were also analyzed using MCC, to evaluate the model globally, based on the results obtained from the confusion matrix.

4.4 Alignment of Protein Sequences

The 3D sequences and structures of GPDH, DHODH, PTR1, TR, and UGPase in *L. major* and *L. braziliensis* were obtained from the GenBank database (National Center for Biotechnology Information, Bethesda MD, USA, https://www.ncbi.nlm.nih.gov/) [50].

Then, a global alignment was performed, using the web tool Clustal Omega (EMBL-EBI, Cambridgeshire, UK, (https://www.ebi.ac.uk/Tools/msa/clustalo/), which aligns all protein sequences entered by a user. The alignment facilitated the investigation of the active site and the determination of the similarity and shared identity among the enzymes between the two species of Leishmania.

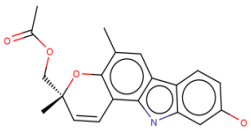
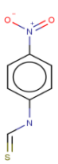
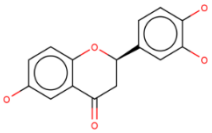
4.5 Homology Modeling

The sequences of the enzymes and species selected in the study were obtained from the GenBank database (National Center for Biotechnology Information, Bethesda MD, USA, https://www.ncbi.nlm.nih.gov/) [50], and the template structures were obtained from the Protein Data Bank (PDB, https://www.rcsb.org/pdb/home/home.do) [51]. Four enzymes were selected for the construction of homology models: GPDH, PTR1, TR, and UGPase. The template enzymes were: GPDH from *L. mexicana* (PDB ID: 1M66), PTR1 from *L. major* (PDB ID: 5L42), TR from L. *infantum* (PDB ID: 5EKB), and UGPAse from *L. major* (PDB ID: 5NZM). The enzyme models were constructed using the homology molecular modeling method in MODELLER 9.20 software (copyright © 1989-2020 Andrej Sali, maintained by Ben Webb at the Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Byers Hall, University of California San Francisco, San Francisco, USA) [52]. Five models were generated, and the lowest energy model was chosen. The stereochemical qualities of the model were assessed by the PSVS webserver (Protein Structure Validation Software suite) (http://psvs-1_5-dev.nesg.org/), using PROCHECK [53]. PROCHECK generates a Ramachandran graph [54], which determines the permitted and disallowed regions of the main chain of amino acids. The structural quality was evaluated using the VERIFY 3D software (saves @2020 - DOE-MBI Services, http://services.mbi.ucla.edu/SAVES/), and the compatibility between the protein sequence and its 3D structure, based on the chemical environment, was analyzed using WHAT IF (http://swift.cmbi.ru.nl/servers/html/index.html).

4.6 Consensus Docking

The consensus docking analysis was performed using four different packages: Molegro Virtual Docker (MVD) (Molexus IVS Rørth Ellevej 3, Odder, Denmark) [55], GOLD 5.6.2 (The Cambridge Crystallographic Data Centre, Cmabridge, USA) [56], AutoDock Vina (Vina) (Molecular Graphics Lab at The Scripps Research Institute) [57], and AutoDock 4.2.6. (AD4) (The Scripps Research Institute, La Jolla, USA), with standard parameters [58]. Five scoring functions were selected for the consensus analysis, including Moldock score (Molexus IVS Rørth Ellevej 3, Odder, Denmark), Goldscore (The Cambridge Crystallographic Data Centre, Cmabridge, USA), ChemPLP (The Cambridge Crystallographic Data Centre, Cmabridge, USA), and the Binding affinity scores in Vina (Molecular Graphics Lab at The Scripps Research Institute) and (The Scripps Research Institute, La Jolla, USA). The enzymes constructed by homology and selected from PDB were used for docking analysis. Information on the enzymes obtained from PDB and their respective inhibitors can be found in Table 13. Initially, all water molecules were removed from the crystalline structure. RMSD ≤ 2.0 Å was used as a criterion for docking success. The consensus strategy consisted of selecting compounds with higher binding affinity prediction values than those for the crystallographic ligands, based on at least three different scoring functions. Then, the values were standardized and averaged. This approach increases the reliability of the fit and increases the number of true positive compounds.

Table 13. Information regarding the selected enzymes deposited in the PDB database and used for docking analysis.

PDB ID	Enzyme	Species	PDB Ligand	Resolution
5NZM	UDP—glucose pyrosphorylase	*L. major*	9ET (Murrayamine-I) 	2.35 Å
4EF9	Dihydroorotate dehydrogenase	*L. major*	4NF 	1.6 Å
5L42	Pteridine reductase 1	*L. major*	6J6 	2.1 Å
4WZH	Dihydroorotate	*L.*	-	2.12 Å

	dehydrogenase	*braziliensis*

4.7 Prediction of Drug Resistance

Genetic variations in the target enzymes were analyzed by searching the TritrypDB database (VEuPathDB, USA, http://tritrypdb.org/tritrypdb/) to identify SNPs in *L. major* and *L. braziliensis*.

After the identification of SNPs, their presence in the region of the active site was investigated, and the most prevalent mutations in the active site or near them were examined, to verify whether these mutations resulted in structural changes or interfered with compound interactions. The mutations were designed using the UCSF Chimera program (Visualization and Informatics – RBVI, San Francisco, USA) [59]. Then, MVD docking was performed, to assess the binding affinity with lignans in the presence of mutations.

4.8 Molecular Dynamics Simulations

MD simulations were performed to estimate the flexibility of interactions between proteins and ligands, using GROMACS 5.0 software (European Union Horizon 2020 Programme, Sweden) [60,61]. The protein and ligand topologies were also prepared using the GROMOS96 54a7 force field. The MD simulation was performed using the SPC water model of point load, extended in a cubic box [62]. The system was neutralized by the addition of ions (Cl^- and Na^+) and minimized, to remove bad contacts between complex molecules and the solvent. The system was also balanced at 300K, using the 100 ps V-rescale algorithm, represented by NVT (constant number of particles, volume, and temperature), followed by equilibrium at 1 atm of pressure, using the Parrinello-Rahman algorithm as the NPT (constant pressure particles and temperature), up to 100 ps. DM simulations were performed in 5,000,000 steps, at 10 ns. To determine the flexibility of the structure and whether the complex is stable close to the experimental structure, RMSD values of all Cα atoms were calculated relative to the starting structures. RMSF values were also analyzed, to understand the roles played by residues near the receptor binding site. The RMSD and RMSF graphs were generated in Grace software (Grace Development Team, http://plasma-gate.weizmann.ac.il/Grace/).

4.9 Isolation and Identification of Lignans

The compounds secoisolariciresinol [63], pinoresinol-4-*O*-β-D-glucopyranoside [64], pinoresinol-4-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside [65], epipinoresinol-4-*O*-β-D-glucopyranoside [66] and were isolated from the fractionation of the crude ethanolic extract of thespecies *Justicia aequilabris* (Nees) Lindau [67] collected in the city of Puxinanã-PB, Brazil, registered in SisGen under the number: A35A42B. The compounds were identified by nuclear magnetic resonance and high-resolution mass spectrometry (HRMS).

Furofuran lignans have their configuration well established in literature, according to Shao et al. [66], the 7, 9′:7′, 9-diepoxi moiety of furofuran lignans of natural origin occurs in the cis-fused configuration. The experiments carried out have demonstrated that chemical deviations of the ΔδH-9 and ΔδH-9′ are resulted of relative configurations of the C-7/C-8 and C-7′/C-8′. Therefore, the authors present the values for chemical displacement differences H2-9 and H2-9′ in different solvents. Thus, compound 158 has the value of ΔδH-9; H-9′ both = 0.40, classifying this molecule as 7-H/8-H trans, 7′-H/8′-H trans. Substance 159 showed ΔδH-9 = 0.34 and ΔδH-9′ = 0.65, determining 7-H/8-H cis and 7′-H/8′-H trans positions. Structure 160 showed Δδ H-9; H-9′ = 0.40, indicating 7-H/8-H and 7′-H/8′-H as trans configuration. All the compounds were analyzed on DMSO-d6. The absolute configuration of compound 156 was deduced by utilizing the experimental and calculated electronic circular dichroism (ECD) data, assigned to the molecule the 8R and 8’R configuration.

4.10 Activity Tests against *L. major* and *L. braziliensis* Axenic Promastigotes

L. major (MHOM/IL/80/Friedlin) and *L. braziliensis* (MHOM/BR/1975/M2903) species were maintained in vitro, as promastigotes, at 26 °C in supplemented Schneider insect medium (20% SFB, 100 U/mL penicillin and 100 µg/mL streptomycin, pH 7), as described by Rodrigues et al. [68]. The growth inhibition assay for the parasites was performed using promastigote forms in the logarithmic phase, which were grown in 96-well plates containing supplemented Schneider insect medium and 1 × 106 parasites/mL, and was performed in triplicate, using different concentrations of lignans (1.56–12.5 µM) and reference meglumine antimoniate drugs (200–40,000 µM) and amphotericin B (0.031–2 µM). The negative control contained neither the reference nor the tested compounds. The culture plates were maintained in a biological oxygen demand incubator (Eletrolab EL202, São Paulo, Brazil), at 26 °C, for axenic promastigotes. After 2 days under these conditions, 10 µL of 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide (MTT, 5 mg/mL) was added to each well, and the cell

culture plates were incubated for 4 h before adding 50 µL of 10% sodium dodecyl sulfate (SDS) solution, to solubilize the formazan crystals. The optical density of the culture was measured in a microplate spectrophotometer reader, at 540 nm (Biosystems ELx800 model, Curitiba, Brazil).

5. CONCLUSIONS

Leishmaniasis is endemic in more than 90 countries, affecting low-income populations. Leishmaniasis is estimated to affect 2 million people annually, worldwide, and more than 20,000 deaths per year are due to complications from the disease. *L. major* and *L. braziliensis* are responsible for CL, which represents more than 90% of cases in several countries. Because current treatments can result in drug-resistance and are often associated with side effects, due to high toxicity index values, lignans have been investigated as a treatment alternative, particularly because many lignans have great ADMET profiles.

We used simple programs and rules to calculate the absorption, bioavailability, pharmacokinetics, and to select lignans with good properties. We were able to screen 33 promising lignans from a set of 160 compounds, which were subjected to several computational and experimental approaches to investigate their leishmanicidal potentials.

The generated predictive models obtained excellent performance results, with an accuracy greater than 81%, and selected 11 lignans with active potential probabilities ranging between 50% and 57%, for *L. major*. For *L. braziliensis*, an accuracy greater than 79% was achieved, and the model selected 21 lignans, with activity probabilities between 50% and 75%.

To investigate the mechanism of action for lignans and to evaluate their selectivity for five enzymes in Leishmania, a consensus docking analysis was performed, to guarantee the reliability of the RF model and to reduce the number of false positives. Eleven lignans were found to be potentially active against *Lm*GPDH, 10 against *Lm*DHODH, 9 against *Lm*PTR1, 9 against *Lm*TR, and 11 against *Lm*UGPase. Among the 21 lignans analyzed in *L. braziliensis*, 20 were potentially active against *Lb*GPDH, 21 against *Lb*DHODH, 19 against *Lb*PTR1, 16 against *Lb*TR, and 15 against *Lb*UGPase.

A combined analysis, based on both ligand and structure, was able to identify potentially active molecules, using both RF and multitarget models, resulting in the identification of 23

potentially active, multitarget molecules against *L. major* and/or *L. braziliensis*, with 10 compounds common to both species.

Due to concerns regarding the development of drug resistance during the treatment of Leishmaniasis, the present study investigated the presence of SNPs, which may contribute to the development of drug resistance. Among the 14 SNPs identified, only four presented polymorphic alleles with relevant frequencies, between 40% and 50%. We also found that none of the amino acids affected by the SNPs were located near the active sites of studied proteins, which reduces the likelihood of developing drug resistance.

MD simulations revealed that most of the studied enzymes are stable under various conditions, including various solvents, ions, temperatures, and pressure, with only small variations observed for some complexed compounds. Therefore, the binding affinity between proteins and ligands is unlikely to be affected by environmental changes. In addition, none of the amino acids responsible for the enzymatic conformational changes were in the active site, except for those in the DHODH-83 complex, which allows the active site to remain stable.

Four lignans with excellent ADMET profiles, which are considered to be potentially active and multitarget inhibitors for the studied enzymes, were isolated from *Justicia aequilabris* (Nees) Lindau [65] and subjected to in vitro tests. Lignans were collected in the city of Puxinanã-PB, Brazil. The results showed that only lignan (**159**) epipinoresinol-4-*O*-β-D-glucopyranoside achieved antileishmanial activity against promastigotes forms of *L. major*, with an IC_{50} value of 36.51 µM. In *L. braziliensis*, compounds (**156**) secoisolariciresinol, (**158**) pinoresinol-4-*O*-β-D-glucopyranoside, (**159**) epipinoresinol-4-*O*-β-D-glucopyranoside, and (**160**) pinoresinol-4-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside inhibited growth, with IC_{50} values of 9.28, 36.35, 5.39, and 13.77, respectively. Compounds (**156**) secoisolariciresinol, (**159**) epipinoresinol-4-*O*-β-D-glucopyranoside, and (**160**) pinoresinol-4-*O*-β-D-apiofuranosyl-(1→2)-β-D-glucopyranoside showed excellent potential as growth inhibitors for the promastigote stage of the parasite. When compared with the values obtained from the biological activity prediction, using the RF models, we noticed that the probability of activity for *L. major* varied 50%–60%, whereas those values for *L. braziliensis* ranged 50%–75%. We suggest that it may be preferable to test compounds with probabilities of activity above 60% to obtain good results and that the computational approach can be used to guide experimental research.

Supplementary Materials: The following are available. Table S1: Lignans and neolignans with good ADMET profiles, Table S2: Predictive assessment of lignin and neolignans toxicity for the evaluated parameters, Figure S3. Alignment of the GPDH protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black

boxes indicate the active site, Figure S4: Alignment of the DHODH protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site, Figure S5: Alignment of the PTR1 protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site, Figure S6: Alignment of the TR protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site, Figure S7: Alignment of the UGPase protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site, Table S: Average of all energy values (EM) obtained from the five scoring functions, for each lignan, and the probability value of potential consensus docking activity (PDC), for each studied enzyme in *L. major*. Absent values indicate the molecules that were eliminated during this evaluation, Table S10: Average of all energy values (EM) obtained from the five scoring functions, for each lignan, and the probability value of potential consensus docking activity (P_DC), for each studied enzyme in *L. braziliensis*. Absent values indicate the molecules that were eliminated during this evaluation.

Author Contributions: Conceptualization, M.T.S., L.S., and F.J.B.M.-J.; methodology, M.d.S.M., F.J.B.M.-J., K.A.d.F.R., J.P.R.eS., T.A.d.L.N., J.M.S.d.S., A.F.M.M., and G.C.S.R., J.F.T.; software, M.T.S. and M.d.S.M.; validation, M.d.S.M., F.J.B.M.-J., and K.A.d.F.R.; formal analysis, M.d.S.M. and G.C.S.R.; investigation, M.d.S.M and J.P.R.eS.; resources, M.T.S.; data curation, M.d.S.M.; writing—Original draft preparation, M.d.S.M., M.T.S., and F.J.B.M.-J.; writing—Review and editing, M.d.S.M., M.T.S., L.S.; F.J.B.M.-J., and K.A.d.F.R.; visualization, M.d.S.M.; supervision, L.S. and M.T.S.; project administration, M.d.S.M.; funding acquisition, M.T.S. and F.J.B.M.-J.

Funding: Coordination of Improvement of Higher Education Personnel—Brazil (CAPES)—Financing Code 001. This work was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant numbers 431254/2018-4, 309648/2019-0 and 308590/2017-1.

Acknowledgments: Coordination of Improvement of Higher Education Personnel—Brazil (CAPES)—Financing Code 001 and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant numbers grant numbers 431254/2018-4, 309648/2019-0 and 308590/2017-1.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Bezerra, J.M.T.; de Araújo, V.E.M.; Barbosa, D.S.; Martins-Melo, F.R.; Werneck, G.L.; Carneiro, M. Burden of leishmaniasis in Brazil and federated units, 1990-2016: Findings from Global Burden of Disease Study 2016. *PLoS Negl. Trop. Dis*. **2018**, *12*, 1–19, doi:10.1371/journal.pntd.0006697.
2. Rath, S.; Augusto Trivelin, L.; Imbrunito, T.R.; Tomazela, D.M.; Jesús, M.N.; Marzal, P.C.; Andrade Júnior, H.F.; Tempone, A.G. Antimoniais empregados no tratamento da leishmaniose: Estado da arte. *Quim. Nova*. **2003**, *26*, 550–555, doi:10.1590/S0100-40422003000400018.
3. Anversa, L.S.; Tiburcio, M.G.S.; Richini-Pereira, V.; Nia, B.; Ramirez, L.E. Human leishmaniasis in Brazil: A general review. *Rev. Assoc. Med. Bras*. **2018**, *64*, 281–289, doi:10.1590/1806-9282.64.03.281.
4. Brasil. Miltefosina Para o Tratamento Da Leishmaniose Tegumentar. *Conitec*. 2018, *27*. Available online: http://conitec.gov.br/images/Relatorios/2018/Relatorio_Miltefosina_LeishmanioseTegumentar.pdf (10/02/2020).
5. Singh, K.; Garg, G.; Ali, V. Current Therapeutics, Their Problems and Thiol Metabolism as Potential Drug Targets in Leishmaniasis. *Curr. Drug Metab*. **2016**, *17*, 897–919, doi:10.2174/1389200217666160819161444.

6. Ferreira, L.L.G.; Andricopulo, A.D. Chemoinformatics strategies for leishmaniasis drug discovery. *Front. Pharmacol*. **2018**, *9*, 1–11, doi:10.3389/fphar.2018.01278.
7. Akhoundi, M.; Kuhls, K.; Cannet, A.; Votypka, J.; Marty, P.; Delaunay, P.; Sereno, D. A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLoS Negl. Trop. Dis*. **2016**, *10*, 1–40, doi:10.1371/journal.pntd.0004349.
8. Balaña-Fouce, R.; Pérez Pertejo, M.Y.; Domínguez-Asenjo, B.; Gutiérrez-Corbo, C.; Reguera, R.M. Walking a tightrope: Drug discovery in visceral leishmaniasis. *Drug Discov. Today* **2019**, *24*, 1209–1216, doi:10.1016/j.drudis.2019.03.007.
9. Akhoundi, M.; Downing, T.; Votýpka, J.; Kuhls, K.; Lukes, J.; Cannet, A.; Ravel, C.; Marty, P.; Delaunay, P.; Kasbari, M.; et al. Leishmania infections: Molecular targets and diagnosis. *Mol. Aspects. Med*. **2017**, *57*, 1–29, doi:10.1016/j.mam.2016.11.012.
10. de Vries, H.J.C.; Reedijk, S.H.; Schallig, H.D.F.H. Cutaneous Leishmaniasis: Recent Developments in Diagnosis and Management. *Am. J. Clin. Dermatol*. **2015**, *16*, 99–109, doi:10.1007/s40257-015-0114-z.
11. Milling, S. Ageing dangerously; homing of senescent CD8 T cells in cutaneous Leishmaniasis. *Immunology* **2020**, *159*, 355–356, doi:10.1111/imm.13188.
12. Gilberto de, S.; William Gustavo, L.; Flávio José dos, S.; Francisco, A.; Macías, J.M.; González, M.; Rafael Gonçalves, T.-N.; João Máximo de Siqueira, E.S. Toxicity and Anti-promastigote Activity of Benzoxazinoid Analogs Against Leishmania (Viannia) braziliensis and Leishmania (Leishmania) infantum. *Adv. Pharm. Bull*. **2020**, *10*, 119–124.
13. Porcino, G.N.; Antinarelli, L.M.R.; Maia, A.C.R.G.; Faria-Pinto, P.; Taunay-Rodriges, A.; Zech Coelho, P.M.; Nelson, D.L.; Penido, M.L.O; Coimbra, E.S.; Vasconcelos, E.G. The alkylaminoalkanethiosulfuric acids exhibit in-vitro antileishmanial activity against Leishmania (Viannia) braziliensis: A new perspective for use of these schistosomicidal agents. *J. Pharm. Pharmacol*. **2019**, *71*, 1784–1791, doi:10.1111/jphp.13163.
14. Santos, A.G.A.; Lima, L.L.; Mota, C.A.; Gois, M.B.; Fernandes, A.C.B.S.; Silveira, T.G.V.; Sant'Ana, D,.M.G.; Nougueira dde Melo, G.A. Insights of Leishmania (Viannia) braziliensis infection in golden hamster (Mesocricetus auratus) intestine. *Biomed. Pharmacother*. **2018**, *106*, 1624–1632, doi:10.1016/j.biopha.2018.07.120.
15. Rêgo, F.D.; Fradico, J.R.B.; Teixeira-Carvalho, A.; Gontijo, C.M.F. Molecular variants of Leishmania (Viannia) braziliensis trigger distinct patterns of cytokines and chemokines expression in golden hamster. *Mol. Immunol*. **2019**, *106*, 36–45, doi:10.1016/j.molimm.2018.12.013.
16. Baneth, G.; Yasur-Landau, D.; Gilad, M.; Nachum-Biala, Y. Canine leishmaniosis caused by Leishmania major and Leishmania tropica: Comparative findings and serology. *Parasites Vectors* **2017**, *10*, 1–9, doi:10.1186/s13071-017-2050-7.
17. Zutshi, S.; Kumar, S.; Sarode, A.; Roy, S.; Sarkar, A.; Saha, B. Leishmania major adenylate kinase immunization offers partial protection to a susceptible host. *Parasite Immunol*. **2020**, *42*, 1–7, doi:10.1111/pim.12688.
18. Iniguez, E.; Schocker, N.S.; Subramaniam, K.; Portillo, S.; Montoya, A.L., Al-Salem, W.S.; Torres, C.L.; Rodrigues, F.; Moreira, O.C.; Acosta-Serrano, A.; et al. An α-Gal-containing neoglycoprotein-based vaccine partially protects against murine cutaneous leishmaniasis caused by Leishmania major. *PLoS Negl. Trop. Dis*. **2017**, *11*, 1–25, doi:10.1371/journal.pntd.0006039.
19. No, J.H. Visceral leishmaniasis: Revisiting current treatments and approaches for future discoveries. *Acta Trop*. **2016**, *155*, 113–123, doi:10.1016/j.actatropica.2015.12.016.

20. Alvarez, V.E.; Niemirowicz, G.T.; Metacaspases, C.J. Autophagins and Metallocarboxypeptidases: Potential New Targets for Chemotherapy of the Trypanosomiases. *Curr. Med. Chem*. **2013**, *20*, 3069–3077.
21. Veras, P.S.T.; Ramos, P.I.P.; De Menezes, J.P.B. In search of biomarkers for pathogenesis and control of leishmaniasis by global analyses of leishmania-infected macrophages. *Front. Cell. Infect. Microbiol*. **2018**, *8*, 1–18, doi:10.3389/fcimb.2018.00326.
22. Borges, A.F.; Gomes, R.S.; Ribeiro-Dias, F. Leishmania (Viannia) guyanensis in tegumentary leishmaniasis. *Pathog. Dis*. **2018**, *76*, 1–10, doi:10.1093/femspd/fty025.
23. Pilkington, L.I. Lignans: A chemometric analysis. *Molecules* **2018**, *23*, 1–24, doi:10.3390/molecules23071666.
24. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem*. **2015**, *8*, 37–47, doi:10.2147/AABC.S70333.
25. Jaghoori, M.M.; Bleijlevens, B.; Olabarriaga, S.D. 1001 Ways to run AutoDock Vina for virtual screening. *J. Comput. Aided. Mol. Des*. **2016**, *30*, 237–249, doi:10.1007/s10822-016-9900-9.
26. Choe, J.; Suresh, S.; Wisedchaisri, G.; Kennedy, K.J.; Gelb, M.H.; Hol, W.G.J. Anomalous differences of light elements in determining precise binding modes of ligands to glycerol-3-phosphate dehydrogenase. *Chem. Biol*. **2002**, *9*, 1189–1197, doi:10.1016/S1074-5521(02)00243-0.
27. Pinto Pinheiro, M.; da Silva Emery, F.; Cristina Nonato, M. Target Sites for the Design of Anti-trypanosomatid Drugs Based on the Structure of Dihydroorotate Dehydrogenase. *Curr. Pharm. Des*. **2013**, *19*, 2615–2627, doi:10.2174/1381612811319140011.
28. Di Pisa, F.; Landi, G.; Dello Iacono, L.; Pozzi, C.; Borsari, C.; Ferrari, S.; Santucci, M.; Santarem, N.; Cordeiro-da-Silva, A.; Moraes, C.B.; et al. Chroman-4-one derivatives targeting pteridine reductase 1 and showing anti-parasitic activity. *Molecules* **2017**, *22*, 1–16, doi:10.3390/molecules22030426.
29. Alves, V.M.; Braga, R.C.; Muratov, E.N.; Horta, C. Quim. Nova,. **2018**, *41*, 202–212.
30. Chang, M.W.; Ayeni, C.; Breuer, S.; Torbett, B.E. Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *PLoS ONE* **2010**, *5*, 1–9, doi:10.1371/journal.pone.0011955.
31. Ren, X.; Shi, Y.S.; Zhang, Y.; Liu, B.; Zhang, L.H.; Peng, Y.B.; Zeng, R. Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *J. Chem. Inf. Model*. **2018**, *58*, 1662–1668, doi:10.1021/acs.jcim.8b00329.
32. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved Protein—Ligand Docking Using GOLD. **2003**, *623*, 609–623.
33. Chaput, L.; Mouawad, L. Efficient conformational sampling and weak scoring in docking programs? Strategy of the wisdom of crowds. *J. Cheminform*. **2017**, *9*, 1–18, doi:10.1186/s13321-017-0227-x.
34. Stevanović, S.; Perdih, A.; Senćanski, M.; Glisic, S.; Duarte, M.; Tomás, A.M.; Sena, F.V.; Sousa, F.M.; Pereira, M.M.; Solmajer, T. In Silico Discovery of a Substituted 6-Methoxy-quinalidine with Leishmanicidal Activity in Leishmania infantum. *Molecules* **2018**, *23*, 1–16, doi:10.3390/molecules23040772.
35. Steiner, T.; Lamerz, A.C.; Hess, P.; Breithaupt, C.; Krapp, S.; Bourenkov, G.; Huber, R.; Gerardy-Schahn, R.; Jacob, U. Open and closed structures of the UDP-glucose pyrophosphorylase from Leishmania major. *J. Biol. Chem*. **2007**, *282*, 13003–13010, doi:10.1074/jbc.M609984200.

36. Chibli, L.A.; Schmidt, T.J.; Nonato, M.C.; Calil, F.A.; Da Costa, F.B. Natural products as inhibitors of Leishmania major dihydroorotate dehydrogenase. *Eur. J. Med. Chem*. **2018**, *157*, 852–866, doi:10.1016/j.ejmech.2018.08.033.
37. Eldehna, W.M.; Almahli, H.; Ibrahim, T.M.; Fares, M.; Al-Warhi, T; Boeckler, F.M.; Bekhit, A.A.; Abbel-Aziz, H.A. Synthesis, in vitro biological evaluation and in silico studies of certain arylnicotinic acids conjugated with aryl (thio)semicarbazides as a novel class of anti-leishmanial agents. *Eur. J. Med. Chem*. **2019**, *179*, 335–346, doi:10.1016/j.ejmech.2019.06.051.
38. da Silva, A.D.; dos Santos, J.A.; Machado, P.A.; Alves, L.A.; Laque, L.C.; de Souza, V.C.; Coimbra, E.S.; Capriles, P.V.S.Z. Insights about resveratrol analogs against trypanothione reductase of Leishmania braziliensis: Molecular modeling, computational docking and in vitro antileishmanial studies. *J. Biomol. Struct. Dyn*. **2019**, *37*, 2960–2969, doi:10.1080/07391102.2018.1502096.
39. Xu, W.H.; Zhao, P.; Wang, M.; Liang, Q. Naturally occurring furofuran lignans: Structural diversity and biological activities. *Nat. Prod. Res*. **2019**, *33*, 1357–1373, doi:10.1080/14786419.2018.1474467.
40. Teponno, R.B.; Kusari, S.; Spiteller, M. Recent Advances in Research on Lignans and Neolignans. *Nat. Prod. Rep*. **2016**, *33*, 1044–1092, doi:10.1039/c6np00021e.
41. Muhit, M.A.; Umehara, K.; Mori-Yasumoto, K.; Noguchi, H. Furofuran Lignan Glucosides with Estrogen-Inhibitory Properties from the Bangladeshi Medicinal Plant Terminalia citrina. *J. Nat. Prod*. **2016**, *79*, 1298–1307, doi:10.1021/acs.jnatprod.5b01042.
42. Mandal, S.; Moudgil, M.; Mandal, S.K. Rational drug design. *Eur. J. Pharmacol*. **2009**, *625*, 90–100, doi:10.1016/j.ejphar.2009.06.065.
43. Bento, A.P.; Gaulton, A.; Hersey, A.; Bellis, L.J.; Chambers, J.; Davies, M.; Kruger, F.A.; Light, Y.; Mak, L.; McGlinchey, S.; et al. The ChEMBL bioactivity database: An update. *Nucleic Acids Res*. **2014**, *42*, 1083–1090, doi:10.1093/nar/gkt1031.
44. Fourches, D.; Muratov, E.; Tropsha, A. Trust, But Verify : On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204.
45. Fourches, D.; Muratov, E.; Tropsha, A. Curation of chemogenomics data. *Nat. Chem. Biol*. **2015**, *11*, 535, doi:10.1038/nchembio.1881.
46. Lorenzo, V.P.; Filho, J.M.B.; Scotti, L.; Scotti, M.T. Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B. *Braz. J. Pharmacogn.* **2015**, *25*, 690–697, doi:10.1016/j.bjp.2015.08.005.
47. Xu, J.; Hagler, A. Chemoinformatics and drug discovery. *Molecules* **2002**, *7*, 566–600, doi:10.3390/70800566.
48. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. DRAGON software: An easy approach to molecular descriptor calculations. *Match* **2006**, *56*, 237–248.
49. Dos Santos Maia, M.; de Sousa, N.F.; Rodrigues, G.C.S.; Monteiro, A.F.M.; Scotti, M.T.; Scotti, L. Lignans and Neolignans anti-tuberculosis identified by QSAR and Molecular Modeling. *Comb. Chem. High Throughput Screen.* **2020**, 1-13, doi:10.2174/1386207323666200226094940.
50. Wishart, D.S. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. **2006**, *34*, D668–D672, doi:10.1093/nar/gkj067.
51. Brice, M.D.; Rodgers, J.R.; Kennard, O. The Protein Data Bank. *Nat. New Biol*. **1977**, *324*, 319–324.
52. Šali, A.; Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol*. **1993**, *234*, 779–815.

53. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **1993**, *26*, 283–291, doi:10.1107/s0021889892009944.
54. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S. Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins* **2003**, *50*, 437–450.
55. Thomsen, R.; Christensen, M.H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem*. **2006**, *49*, 3315–3321, doi:10.1021/jm051197e.
56. Sargis, D.; Arthur, J. Small-Molecule Library Screening by Docking with PyRx. *Chem. Biol*. **2014**, *1263*, 243–250.
57. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem*. **2009**, *30*, 2785–2791, doi:10.1002/jcc.21256.
58. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem*. **2004**, *25*, 1605–1612, doi:10.1002/jcc.20084.
59. Drunen A links open overlay panelH. JCB va. der S va. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun*. **1995**, *43*, 1–3.
60. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Berk, H.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25, doi:10.1016/j.softx.2015.06.001.
61. Bondi, A. Van der Waals Volumes and Radii. *J. Phys*. **1964**, *68*, 441–451.
62. Wang, X.-P.; Song, G.-Y.; Chen, W.-H.C. Chemical Constituents from Roots of Millettia speciose. *Chin. Herb Med*. **2016**, *8*, 385–389.
63. Huang, S.W.; Qiao, J.W.; Sun, X.; Gao, P.Y.; Li, L.Z.; Liu, Q.B.; Sun, B.; Wu, D.; Song, S.J. Secoiridoids and lignans from the leaves of Diospyros kaki Thunb. with antioxidant and neuroprotective activities. *J. Funct. Foods* **2016**, *24*, 183–195, doi:10.1016/j.jff.2016.03.025.
64. Abe, F.; Yamauchi, T. Lignan glycosides from Parsonsia laevigata. *Phytochemistry* **1989**, *28*, 1737–1741, doi:10.1016/S0031-9422(00)97835-X.
65. Shao, S.Y.; Yang, Y.N.; Feng, Z.M.; Jiang, J.S.; Zhang, P.C. An Efficient Method for Determining the Relative Configuration of Furofuran Lignans by 1 H NMR Spectroscopy. *J. Nat. Prod*. **2018**, *81*, 1023–1028, doi:10.1021/acs.jnatprod.8b00044.
66. Corrêa, G.M.; Alcântara, A.F.D.C. Chemical constituents and biological activities of species of Justicia—A review. *Revista Brasileira De Farmacognosia* **2012**, *22*, 220–238.
67. Rodrigues, K.A.D.F.; Dias, C.N.D.S.; Neris, P.L.D.N.; Rocha, Jda.C.; Scotti, M.T.; Scotti, L.; Mascarenhas, S.R.; Veras, R.C.; de Medeiros, I.A.; Keesen, Tde.S.; et al. 2-Amino-thiophene derivatives present antileishmanial activity mediated by apoptosis and immunomodulation in vitro. *Eur. J. Med. Chem*. **2015**, *106*, 1–14, doi:10.1016/j.ejmech.2015.10.011.

Sample Availability: Samples of the compounds ..... are available from the authors.

Capítulo 3

Neste capítulo, 47 neolignanas foram selecionadas do banco de dados ChEMBL e submetidas a uma triagem virtual baseada no ligante e na estrutura, com análise consenso entre os dois métodos. Para a análise de QSAR, foram construídos modelos preditivos a partir de dados de atividades biológicas de estruturas químicas testadas contra importantes alvos do *Trypanossoma cruzi*. Os modelos das enzimas cruzain e CYP51 foram obtidos e contribuíram para predizer a atividade biológica de neolignanas para esses alvos. Além disso, análises de acoplamento consenso nas enzimas cruzain, CYP51 e TR foram realizadas.

As neolignanas também foram submetidas à propriedades farmacocinéticas que avaliaram o perfil ADMET das estruturas químicas innvestigadas. Em seguida, após a aplicação das metodologias de triagem virtual e ADMET, foram realizadas simulações de dinâmica molecular com os dois compostos que foram possíveis isolar para avaliar a flexibilidade do alvo em que as neolignanas foram mais seletivas. Cálculos de energia de livre contribuíram para complementar os estudos de docking e dinâmica molecular. Por fim, as lignanas isoladas foram testadas contra as formas epimastigotas e tripomastigotas de *T. cruzi* para validar a abordagem *in silico* utilizada.

Virtual screening based on ligand and structure with in vitro assessment of neolignans against *Trypanossoma cruzi*

Mayara dos Santos Maia[1], Rodrigo Silva de Andrade[2], Julyanne Maria Saraiva de Sousa[3], Gabriela Cristina Soares Rodrigues[1], Marcelo Sobral da Silva[2], Josean Fechine Tavares[2], Klinger Antonio da Franca Rodrigues[3], Luciana Scotti[1] and Marcus Tullius Scotti[1*.]

[1]Laboratory of Cheminformatics, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, 58051-900, João Pessoa-PB, Brazil; mayarasmaia@hotmail.com; gaby.ecologia@gnail.com; luciana.scotti@gmail.com; mtscotti@gmail.com.
[2]Multi-User Characterization and Analysis Laboratory, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, 58051-900, João Pessoa-PB, Brazil; rodrigo@ltf.ufpb.br; marcelosobral.ufpb@gmail.com; josean@ltf.ufpb.br
[3]Infectious Diseases Laboratory, Federal University of Parnaíba Delta, São Benedito, 64202-020, Parnaíba-PI, Brazil; jully.yanne@gmail.com; klinger.antonio@gmail.com.

*Correspondence: mtscotti@gmail.com

ORCID
Mayara dos Santos Maia: 0000-0002-5306-8362
Rodrigo Silva Andrade: 0000-0002-3378-6500
Julyanne Maria Saraiva de Sousa: 0000-0001-9041-0860
Gabriela Cristina Soares Rodrigues: 0000-0002-9969-0325
Marcelo Sobral da Silva: 0000-0003-3451-8468
Josean Fechine Tavares: 0000-0003-0293-2605
Klinger Antonio da Franca Rodrigues: 0000-0003-3904-3529
Luciana Scotti: 0000-0003-1866-4107
Marcus Tullius Scotti: 0000-0003-4863-8057

Artigo submetido na revista: Journal Molecular Sctruture

Fator de impacto: 3.196

ABSTRACT: Chagas disease, caused by the parasite *Trypanosoma cruzi*, occurs most commonly in Latin America. As the treatment is highly toxic and ineffective in the chronic phase of the disease, alternative treatments are needed. We predicted the trypanomicidal potential of 47 neolignans using predictive models of their biological activity and molecular docking. A combined analysis allowed for the selection of potent inhibitors against *Trypanosoma cruzi*. Of these compounds, two were isolated and shown to inhibit the growth of epimastigotes at concentrations of 9.64 and 8.72 µM, and trypomastigote forms at 4.88 and 2.73 µM. Therefore, the compounds (2R, 3R)-2,3-Dihydro-2-(4-methoxyphenyl)-3-methyl-5-(E)-propenylbenzofuran (**46**) and ottomentosa (**47**) could be excellent growth inhibitors for the parasite stages and warrant additional study.

Keywords: Chagas disease; *Trypanossoma cruzi*; neolignan; virtual screening; molecular docking.

1. INTRODUCTION

Trypanosoma cruzi the protozoan parasite that causes Chagas disease (CD), was initially endemic to Latin America, but has spread to other locations such as Canada, the United States, Europe, Australia, and Japan [1]. It currently affects 6−7 million people worldwide and causes approximately 50,000 deaths per year [1,2]. The transmission of *T. cruzi* can occur congenitally, through organ transplantation, blood or, ingestion of food and drinks contaminated by the parasite [3].

The parasites are transmitted predominantly to humans as metacyclic trypomastigote (MT) forms through the contaminated feces of blood-sucking triatomines at the bite site. After internalization by host cells near the entry site, MTs initially reside in a vesicle containing the parasite, the parasitophore vacuole, from where they escape into the host cell's cytoplasm and differentiate in the proliferative amastigote form. After several cycles of replication, amastigotes differentiate into mobile flagellated trypomastigotes, which are released into the bloodstream, from where they can spread by infecting distant tissues or are captured by the triatomine vector during a blood meal. Ingested blood trypomastigotes become epimastigotes in the vector's midgut, multiplying and then differentiating into infectious metacyclic trypomastigotes. Infectious trypomastigotes and intracellular replicative amastigotes are the parasite's clinically relevant life cycle stages targeted for that are targets for drug intervention [3],[4].

Currently, no vaccines prevent diseases caused by trypomastigotas, and although chemotherapeutic drugs have been available for decades, they are highly toxic and have

unpleasant side effects. Benznidazole, a derivative of nitroimidazole, and nifurtimox, a nitrofuran compound, both developed more than 40 years ago, are currently the only drugs available for the treatment of CD [5]. Although, benznidazole is the first-line drug due to its better tolerability, but both drugs have significant side effects [6]. Therefore, patients should be monitored frequently. Unfortunately, the medications available are effective only in the acute phase and in 20% of cases the treatment must be stopped due to side effects [7].

As neolignans are a diverse group of chemical structures present in several plant families and responsible for multiple biological activities, we chose to investigate the trypanomicidal potential of a set of neolignans. We explored computational and experimental studies to select the most promising tripanomicidal compounds that may be effective in the acute and chronic phases of CD and present low toxicity. We identified three important targets (the enzymes cruzain, trupanothione reductase [TR], and sterol 14-alpha demethylase [CYP51]) for the proliferation and survival of the parasite in the parasite's three cellular forms. Thus, the identified compounds may be effective as insecticidal and drug agents in the acute and chronic phases of the disease.

2. MATERIAL AND METHODS

2.1 Data collection and curation

The biological activity and 3D structure data of the enzymes cruzain, trypanothione reductase (TR), and sterol 14-alpha demethylase (CYP51) were investigated. Enzymes data (Table 1) were downloaded from the ChEMBL database (https://www.ebi.ac.uk/chembl/) [8,9] with the codes CHEMBL3563 (cruzain), CHEMBL5131 (TR) and CHEMBL1075110 (CYP51). These compounds were used to build predictive models. The compounds were classified based on the pIC_{50} (-log IC_{50} [mol/l]). The IC_{50} value represents the concentration required for 50% inhibition of the parasites. In addition, 47 neolignans obtained from ChEMBL were evaluated by virtual screening to identify molecules with potential activity against the three main enzymes listed above that are involved in the proliferation and survival of *T. cruzi*, according to the workflows presented by Fourches *et al*. [10]. Three-dimensional structures were generated and standardized using the Standardizer v.18.17.0, (www.chemaxon.org).

Figure 1 shows a schematic depicting all of the methodologies used in this study.

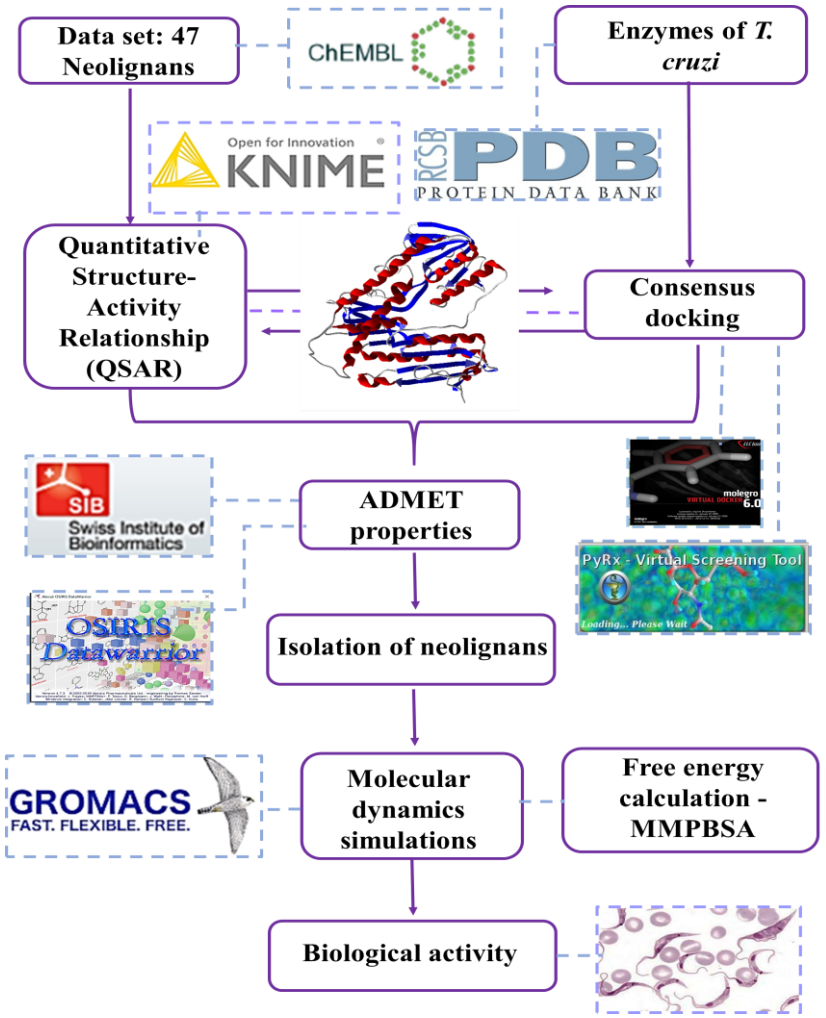


Figure 1. Outline of all study procedures.

Table 1. Set of molecules from the ChEMBL Databases for each cruzain, TR and CYP51 database of *T. cruzi*.

Database	Active molecules	Inactive molecules	Total
Cruzain	217 ($pIC_{50} \geq 5.02$)	255 ($pIC_{50} \leq 5.0$)	472
TR	117 ($pIC_{50} \geq 4.5$)	150 ($pIC_{50} \leq 4.46$)	267
CYP45	271 ($pIC_{50} \geq 4.6$)	317 ($pIC_{50} \leq 4.5$)	588

2.2 Quantitative structure-activity relationship modeling (QSAR)

Knime 3.6.2 software (Knime 3.6.2, Konstanz Information Miner, Zurich, Switzerland, www.knime.org) was used to perform QSAR modeling. Given the success of our previous studies [11-14], we opted to perform a QSAR 3D analysis. To generate descriptors,

all compounds with a solved chemical structure were saved in SDF format and imported into Dragon 7.0 software (Kode Chemioinformatics SRL, Pisa, Italy) [14]. The descriptors generated in Dragon were imported into the Knime software and the random forest algorithm (RF) was used to build prediction models. The applicability domain was estimated, according to procedures previously described [15]. The external performances of the selected models were analyzed for sensitivity (true-positive rate, or active rate), specificity (true-negative rate, or inactive rate), and accuracy (general predictability). Cross-validation was performed to estimate the predictive power of the developed models. In addition, the performance of external models was assessed by ROC analysis. In addition, MCC was used to evaluate the models globally, based on the results obtained from the confusion matrix.

2.3 Consensus docking

Molecular docking was used to investigate the mechanism of action of selected compounds against the enzymes cruzain, TR and CYP51. The 3D structures of the enzymes were obtained from the Protein Data Bank (PDB) (https://www.rcsb.org/pdb/home/home.do) [16]. Information about complexed enzymes and inhibitors can be seen in Table 2. Initially, all water molecules were removed from the crystalline structure and the mean square deviation (RMSD) was calculated from the positions, indicating the degree of reliability of the adjustment. The RMSD provides the connection mode close to the experimental structure and is considered successful if the value is less than 2.0 Å. A consensus analysis using three different scoring functions was used to decrease the number of false positives. The parameters of each docking program used are described below.

Molegro Virtual Docker (MVD) 6.0

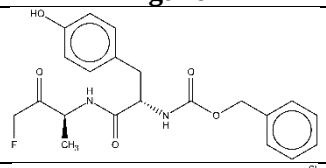
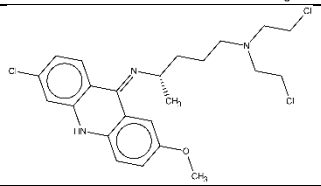
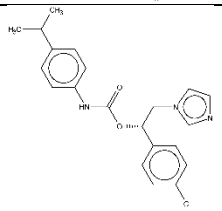
We used the Molegro Virtual Docker v.6.0.1 (MVD) software [17] with its predefined parameters. Then, a docking wizard was created in which the enzymes and ligands were inserted to analyze system's stability through the interactions associated with the enzyme's active site, using the energy value of the MolDock Score [17] as a reference. The MolDock SE (Simplex Evolution) algorithm is based on differential evolution and was used with the following parameters: a total of 10 runs with a maximum of 1,500 iterations using a population of 50 individuals, 2,000 minimization steps for each flexible residue and 2,000 steps of global minimization per run. The MolDock Score (GRID) and PLANTS score

(GRID) scoring function were used to calculate the fit energy values. A GRID was set at 0.3 A and the search sphere radius was set at 15 Å. For the analysis of the ligand energy, electrostatic interactions, hydrogen bonds and sp2-sp2 torsions were evaluated.

AutoDock Vina (Vina)

We used AutoDock Vina [18] under the graphical interface of the PyRx Virtual Screening program tool [19], maintaining the default parameters of the software. This program is based on the genetic algorithm and the empirical energy of the field strength. The protein and ligand files were converted to pdbqt format, and the generated GRID was conducted to the active site region. AutoDock Vina generated 10 conformations for each binder which were used for analysis. The binding affinity scoring function (Kcal/mol) corresponds to the sum of intermolecular and intramolecular contributions and the potentials are based on knowledge and empirical scores.

Table 2. Inhibitor-enzyme complex data for *T. cruzi* enzymes cruzain, TR, and CYP51.

Enzyme	PDB ID	Ligand	Resolution
Cruzain	1AIM		2.0 Å
TR	1GXF		2.7 Å
CYP51	4CK9		2.7 Å

2.4 Prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties

ADME parameters were calculated using the SwissADME open-access web tool (http://www.swissadme.ch) [20], which offers a set of rapid predictive models for the assessment of physicochemical, pharmacokinetic and pharmacological properties. The toxicity prediction was performed in the OSIRIS Property Explorer (https://www.organic-

chemistry.org/prog/peo/) [21], based on the following parameters: mutagenicity, tumorigenicity, reproductive effects and irritability. For absorption, factors included membrane permeability, intestinal absorption and the glycoprotein P substrate or inhibitor. Thus, we investigated compounds that did not exceed more than two violations of the Lipinski rule and for which the logP consensus was not greater than 4.15. In addition, the compounds were not substrates for the permeability glycoprotein enzyme (P-gp). The distribution was evaluated by factors that included the blood-brain barrier (logBB) and the permeability of the CNS. Metabolism was predicted based on the CYP substrate or inhibition models (CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4).

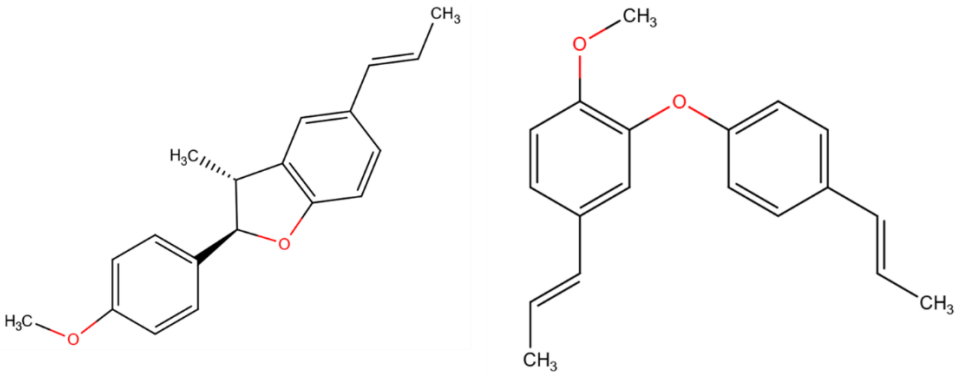
2.5 Isolation and identification of neolignans

Among the neolignans investigated in this study, two were possible to isolate, (2R, 3R)-2,3Di-hydro-2-(4-methoxyphenyl)-3-methyl-5-(E)-propenylbenzofuran (compound **46**) and ottomentosa (compound **47**). The neolignans were isolated from the roots of *Krameria tomentosa* (2.2 kg) collected in the city of Santa Rita-PB, Brazil (7°09'15.4" S, 35°00'14.9" W) in June 2014. Their access registrations in the National Management System of Genetic Patrimony and Associated Traditinal Knowledge (SISGEN) were obtained under number A837A80.

The roots were dried, pulverized and extracted with 95% EtOH at room temperature, and 380 g of the ethanolic extract. A portion (50 g) was suspended in MeOH:H_2O (7:3) and sequentially partitioned with hexane, dichloromethane, ethyl acetate (EtOAc) and *n*-butyl alcohol to obtain the respective extracts. The dichloromethane extract (2 g) was fractioned by medium pressure liquid chromatography packed with silica gel, to obtain 14 fractions. Fraction 1 (20 mg) was subjected to a preparative HPLC to yield compounds **46** (2*R*,3*R*)-2,3-dihydro-2-(4-methoxyphenyl)-3-methyl-5-[1-(*E*)-propenyl]benzofuran (2.1 mg) [22] and **47** (2.4 mg) [23].

The compounds were identified by 1D- and 2D-NMR, high-resolution mass spectrometry (HRMS), optical rotation analysis and by with examination the literature [253][254][255].

Figure 2. Isolated structures of the roots of *Krameria tomentosa*.



2.6 Molecular dynamics simulations

MD simulations were performed to estimate the flexibility of interactions between proteins and ligands, using GROMACS 5.0 software (European Union Horizon 2020 Programme, Sweden) [25, 26]. The protein and ligand topologies were also prepared using the GROMOS96 54a7 force field. The MD simulation was performed using the SPC water model of point load, extended in a cubic box [27]. The system was neutralized by adding ions (Cl^- and Na^+) and minimized, to remove bad contacts between complex molecules and the solvent. The system was also balanced at 300K, using the 100 ps V-rescale algorithm, represented by NVT (constant number of particles, volume, and temperature), followed by equilibrium at 1 atm of pressure, using the Parrinello-Rahman algorithm as the NPT (constant particles pressure and temperature), up to 100 ps. DM simulations were performed in 5,000,000 steps, at 10 ns. To determine the flexibility of the structure and whether the complex is stable and close to the experimental structure, RMSD values of all Cα atoms were calculated relative to the starting structures. RMSF values were also analyzed, to understand the roles played by residues near the receptor binding site. The RMSD and RMSF graphs were generated in Grace software (Grace Development Team, http://plasma-gate.weizmann.ac.il/Grace/) and the protein and ligands were visualized in UCSF Chimera [258]

2.7 Free energy calculations

The Molecular Mechanics - Poisson Boltzmann Surface Area approach (MM/PBSA) was used to calculate the free binding energy of the protein-binding complex in the study of the molecular behavior of the enzymes and its respective ligands. The GROMACS g_mmpbsa

module [29, 30] was applied to estimate the bond-free energy of the selected complex using the trajectory files obtained in the molecular dynamics simulation. The GROMACS MM-PBSA calculation consisted of three steps. First, the potential energy in the vacuum was calculated, and then; the energies of polar and, finally, nonpolar solvation were estimated. The non-polar solvation energy was calculated using the solvent accessible surface area model (SASA). The required input files and solvation energy values were then selected to evaluate the following energetic components: van der Waals energy, electrostatic energy, polar energy of solvation, non-polar solvation energy, and free energy of bonding.

2.8 Activity of selected neolignans against *T. cruzi* epimastigotes and trypomastigotes

Parasites and cells

Epimastigote forms of *T. cruzi*, strain Y, were cultured liver infusion tryptose medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin 10,000 IU/10 mg and kept at 28 ºC in a of biochemical oxygen demand incubator. The parasites used in the experiments were aliquoted from cultures in an exponential growth phase, determined using a 10-day growth curve.

Rhesus monkey renal cells LLC-MK2 (ATCC CCL-7) were cultured in DMEM medium, pH 7.2, supplemented (10% FBS and 1% 100 U/mL penicillin and 100 µg/mL streptomycin) and incubated at 37 °C and 5% CO_2. Maintenance was performed every two days or when the cells reached the confluence.

Trypomastigote forms were obtained by infection of LLC-MK2 cells. The cells were cultured (2x105) in DMEM medium supplemented with 10% FBS and 1% penicillin-streptomycin 10,000 IU/10 mg maintained at 37 ºC and 5% CO_2. After reaching the confluence state, they were infected with a suspension of epimastigotes (1 x 107). The infected cells were kept in DMEM medium with 2% SFB at 37 ºC and 5% CO_2 for six days. Finally, the trypomastigote forms were removed from the supernatant, centrifuged and available for testing [31].

Activity epimastigotes and trypomastigotes

Epimastigote and trypomastigote forms were seeded in 96-well plates (1 x 106) containing different concentrations of substances and LIT medium with 10% SFB for

epimastigote forms and DMEM medium with 2% SFB for trypomastigote forms. Then they were incubated for 48 hours at 28 ºC (epimastigotes) and 37 ºC (trypomastigotes). Growth inhibition was then assessed by quantification in a Neubauer chamber. The positive control was the standard drug benznidazole [32].

Cytotoxicity

Cytotoxicity was evaluated in 96-well plates using the MTT colorimetric assay. Approximately 1 x 106 LLC-MK2 cells per well were incubated in 100 µL of DMEM medium supplemented with FBS at 37 °C and 5% CO_2 for 4 h to determine cell adhesion. First, non-adherent cells were removed by washing with pure DMEM medium. Next, a new supplemented DMEM medium containing increasing concentrations of the substances was added and then incubated at 37 ºC with 5% CO2 for 48 h. After incubation, cytotoxicity was assessed by adding 10 µL of MTT (5 mg/mL). Next, the supernatant was discarded, and the formazan crystals were dissolved by adding 100 µL of DMSO. Finally, the absorbance at 540 nm was measured using an ELISA plate reader [32].

3. RESULTS AND DISCUSSION

3.1 QSAR Modeling

Three prediction models were built using the random forest (RF) algorithm to perform ligand-based virtual screening. For the construction of these models, molecular descriptors were calculated for molecules with known activity against cruzain, TR and CYP51 of *T. cruzi* obtained through the ChEMBL database.

RF models were evaluated for their predictive powers, using the parameters of specificity, sensitivity, precision, accuracy (AUC), positive predicted value (PPV) and negative predicted value (NPV), in addition to performance and robustness, using the receiver's operating characteristic (ROC) curve and Mathews correlation coefficient (MCC). Table 3 describes the characteristics of the models, in terms of predictive power and robustness, and Figure 3 shows the performance of the models. The results showed that the models provided excellent classification, performance, and robustness, except for the CYP51

enzyme, for which the accuracy and specificity values were below 0.6. Therefore, this model was disregarded for the prediction.

After the models were validated, they were used to analyze the set of neolignans for activity against *T. cruzi*. The neolignan bank was then screened to select compounds that are potentially active against cruzain and TR.

The RF model was able to select a compound with active potential, with probabilities from 50% for cruzain (Table 4). The TR model was able to classify all 47 compounds as potentially active, with probabilities ranging between 54% and 84% (Table 4). According to these results, neolignans are potentially active and selective for the TR enzyme.

Table 3. Summary of parameters corresponding to the results obtained for all models.

Enzyme	Validation	Specificity	Sensitivity	Accuracy	PPV	NPV	MCC	AUC
Cruzain	Test	0.84	0.77	0.81	0.81	0.81	0.61	0.86
	Cross	0.80	0.68	0.75	0.75	0.75	0.49	0.82
TR	Test	0.63	0.95	0.85	0.85	0.87	0.65	0.94
	Cross	0.70	0.92	0.85	0.86	0.81	0.65	0.89

Figure 3. Receiver operating characteristic (ROC) curve-generated random forest (RF) model. (1A) Test and (1B) cross-validation for the enzyme cruzain and (2A) test and (2B) cross-validation for the TR.

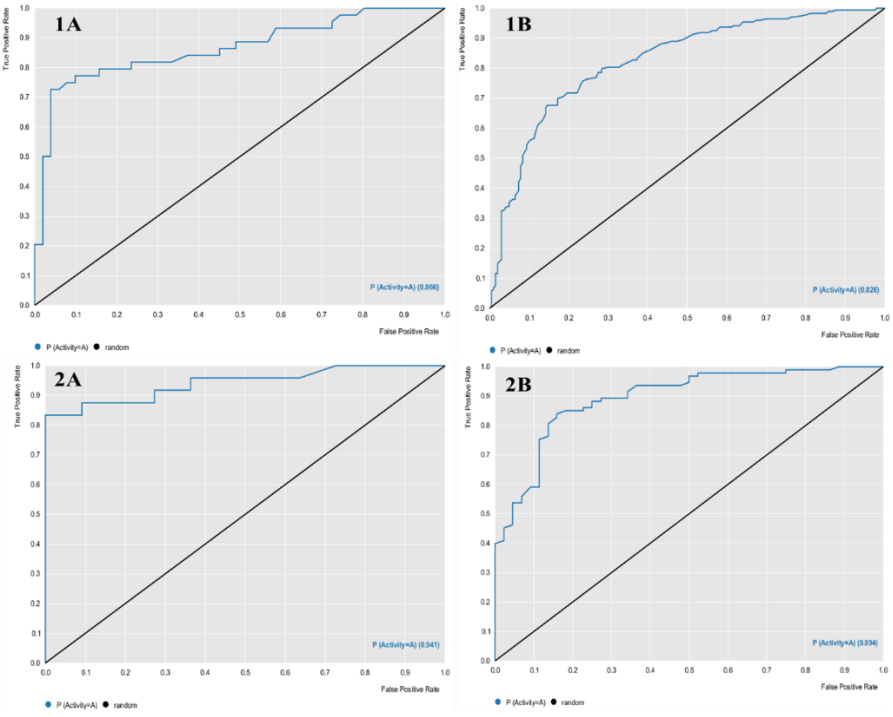


Table 4. Neolignans activity probabilities ($P_{Activity}$) against cruzain and TR as assessed by the RF model. The compounds considered active in the models are highlighted in bold.

ID	pActivity		ID	pActivity		ID	pActivity	
	Cruzain	TR		Cruzain	TR		Cruzain	TR
1	0.37	**0.73**	17	0.38	**0.73**	33	0.46	**0.83**
2	0.38	**0.67**	18	0.27	**0.65**	34	0.31	**0.78**
3	0.40	**0.76**	19	0.41	**0.75**	35	0.48	**0.82**
4	0.39	**0.73**	20	0.40	**0.73**	36	0.41	**0.74**
5	0.29	**0.75**	21	0.31	**0.78**	37	0.34	**0.69**
6	0.35	**0.66**	22	0.36	**0.75**	38	0.35	**0.68**
7	0.31	**0.72**	23	0.29	**0.75**	39	0.34	**0.72**
8	0.48	**0.79**	24	0.29	**0.68**	40	0.37	**0.69**
9	0.41	**0.76**	25	0.39	**0.78**	41	0.37	**0.73**
10	0.37	**0.54**	26	0.30	**0.79**	42	0.34	**0.78**
11	0.35	**0.64**	27	0.48	**0.79**	43	0.47	**0.82**
12	**0.50**	**0.78**	28	0.29	**0.78**	44	0.29	**0.80**
13	0.34	**0.70**	29	0.35	**0.76**	45	0.35	**0.74**
14	0.37	**0.76**	30	0.47	**0.84**	46	0.36	**0.62**
15	0.31	**0.75**	31	0.28	**0.80**	47	0.39	**0.72**
16	0.39	**0.71**	32	0.32	**0.60**			

3.2 Docking consensus

The 47 neolignans also underwent a docking consensus assessment to increase the method's reliability and decrease false positives. The enzymes cruzain, TR and CYP51 were used for docking studies. The docking results were generated using three different scoring functions and validated by redocking the PDB ligand for each enzyme. More negative values indicated better predictions for most scoring functions.

After docking, the results were standardized for each scoring function using the docking probability formula ($Prob_{Dc}$):

$$ProbDc = \frac{E_{Lig}}{E_{MLig}}, se\ E_{Lig} < E_{Inib}$$

where E_{Lig} is the energy of the ligand, E_{MLig} is the energy of the ligand with the highest score, and E_{Inib} is the energy of the inhibitor obtained from the crystallographic data of the test protein. The highest value in $Prob_{Dc}$ is equal to 1. Thus, only those compounds with energy values equal to or greater than the interaction with energy of the crystallographic inhibitor ligand were considered to be potentially active. Then, an average of the standardized results for each scoring function was determined.

The docking results were validated by re-docking the crystallographic ligand and the RMSD of the positions. Re-docking consists of positioning and predicting the binding affinity

of the crystallographic ligand in the region of the enzyme's active site. The RMSD compares and calculates the mean square root deviation of the postures obtained by re-docking and the ligand structure obtained experimentally. For the adjustment to be reliable, the RMSD value must be 2.0 Å or less. The results showed that the targets cruzain, TR and CYP51 obtained 0.77, 0.64 and 0.31 Å, respectively.

Among the 47 neolignans analyzed by molecular docking, two compounds were potentially active against Cruzain. The cruzain test inhibitor had a $Prob_{Dc}$ value of 0.88 and only two neolignans showed values higher than the inhibitor, with values of 0.91 (compound 1) and 0.90 (compound 35). For the TR enzyme, 31 neoligans were considered active, with Pro_{Dc} values ranging from 0.49 to 0.81. The TR inhibitor had a $Prob_{Dc}$ value equal to 0.49. For the CYP51 enzyme, the inhibitor obtained a value of 0.56 and 18 neolignans obtained an equal or greater value. These results indicated that neolignans, in general, were more likely to activate TR and CYP51 proteins, and are not selective for the cruzain enzyme. These results can be seen in Tables 3−5.

3.3 Structure and ligand-based combined analyses

A second consensus analysis was carried out to identify potential lignans and multitarget, which demonstrate the probabilities of being active potentials for more than one protein, based on the RF and docking model. In this case, we used all the results of the prediction of the biological activity of the neolignans and combined them with the docking results. In addition to selecting the active compounds, this combined analysis allowed the selection of the most potent compounds by combining two important methodologies, based on the ligand and structure. For this analysis, the following formula was used:

$$Prob_{Comb} = \frac{(ProbDc + (1 + ESP) \times P_{Activity}}{2 + ESP}, if\ Prob_{Comb} > 0.5$$

where $Prob_{Dc}$ is the probability of a compound being active in the molecular coupling analysis, ESP is the value of the specificity of the RF model and $P_{Activity}$ is the probability value of a compound to be active in the RF model. The combined probability ($Prob_{comb}$), based on the ligand and structure, can increase the predictive power of the models and decrease the number of false positives. For the molecules to be considered potentially active, the probability values must were required to be equal to or greater than 0.5. The higher the Pro_{bcomb} value, the greater the potential of the molecule. The combined probability values

were calculated for all neolignans and each target enzyme, and we analyzed which molecules were multitargets. Only for the CYP51 enzyme was it impossible to calculate the Probcomb value, as we did not have biological activity data for this enzyme due to the low quality of the generated RF model. Therefore, we considered only the ProbDc values for this target. In this case, the $Prob_{comb}$ value needed to be higher than the crystallographic ligand, which was 0.56.

After performing the combined analysis, based on the ligand and structure, and using the formula to identify potentially active and multitarget molecules, we identified 29 potentially active molecules for more than two target enzymes from the entire set of neolignans analyzed. In addition, after the combined probability analysis, we selected the multitarget compounds that passed the applicability domain for all enzymes in this study. Using $Prob_{comb}$, we were able to select 22 compounds with a probability of activity ranging from 50%−65% for cruzain (Table 5). The combined analysis allowed us to select more active compounds for this target than using the $Prob_{Activity}$ and $Prob_{Dc}$ results separately. It was also possible to select 46 neolignans potentially active against the TR enzyme with a $Prob_{comb}$ value ranging from 52−82% (Table 6), while 18 neolignans obtained a probability ranging between 56−88% for the CYP51 enzyme (Table 7).

3.4 Prediction of ADMET properties

The set of 47 neolignans was submitted to several predictive parameters to identify the compounds with the best pharmacokinetic, pharmaco-chemical and pharmacological profiles. Initially, through physicochemical properties, we sought to verify compounds with good absorption, using the lipid rule as a parameter.

We evaluated the absorption and bioavailability properties using the Lipinski rule [33], comprising molecules with molecular weights below 500 Da, calculated LogP (CLogP) values less than five, less than five hydrogen bonding donors, no more than ten hydrogen bonding acceptors and ≤ 10 rotating bonds with excellent absorption and bioavailability. Molecules that violate two or more of these rules do not demonstrate sufficient absorption. We observed that only one neolignan from our set of compounds did not meet this requirement. Therefore, 97.87% of the neolignans showed good absorption and bioavailability.

Factors such as lipophilicity and solubility contribute to the distribution of the drug in vivo, which is a requirement for advancing to preclinical and clinical testings. The most common descriptor for lipophilicity is the partition coefficient between n-octanol: water (log P). The results showed that all neolignans had ideal logP values below 5.0.

Pharmacokinetics are essential for understanding drug metabolism in the body, half-lives, and toxic metabolites. Unfortunately, many compounds fail in the pre-clinical and clinical testing phase due to the effects of metabolism and malabsorption on the brain. Therefore, an early assessment of this effect is necessary and an in silico approach contributes substantially to mitigate adverse reactions that some may experience. The results showed that 47.55% of the neolignans were not a substrate for the CYP enzyme and did not cross the blood-brain barrier.

Toxicity was also evaluated, and we observed that only nine neolignans showed low toxicity in at least one parameter evaluated, such as mutagenicity or tumorigenesis, negative effects on the reproductive system, and irritability. Therefore, 80.85% of neolignans were considered to have the best ADMET properties, as they do not present toxicity risks. Table S1 shows the ADMET profile of the entire set of neolignans.

Table 5. Improved results in the combined probability between the prediction model and molecular docking analysis ($Prob_{comb}$) for potential activity against cruzain. Shown are the compounds considered active (with $Prob_{comb}$ equal to or above 0.50) withvalues of binding energy, molecular docking probability (Pro_{Dc}) and probability of biological activity ($Prob_{Ac}$).

ID	Moldockscore	Pro_{Dc}	Plantscore	Pro_{Dc}	Vina	Pro_{Dc}	Mean ProbDc	ProbAc	$Prob_{Comb}$
1	-116.3	0.94	-100.3	0.90	-5.9	0.89	0.91	0.37	0.56
2	-89.40	0.72	-65.45	0.58	-5.7	0.86	0.72	0.38	0.50
3	-95.33	0.77	-83.19	0.74	-5.2	0.78	0.76	0.40	0.53
4	-88.76	0.71	-70.87	0.63	-5.3	0.80	0.71	0.39	0.50
8	-103.2	0.83	-74.23	0.66	-5.7	0.86	0.78	0.48	0.59
10	-107.3	0.86	-88.76	0.79	-5.7	0.86	0.84	0.37	0.53
12	-83.28	0.67	-37.54	0.33	-5.6	0.84	0.61	0.5	0.54
13	-109.6	0.88	-82.69	0.74	-5.8	0.87	0.83	0.34	0.51
19	-82.13	0.66	-75.93	0.68	-5.6	0.84	0.73	0.41	0.52
20	-96.09	0.77	-81.80	0.73	-5.3	0.80	0.77	0.40	0.53
21	-99.46	0.80	-77.27	0.69	-5.2	0.78	0.76	0.36	0.50
25	-96.83	0.78	-76.85	0.69	-.53	0.80	0.75	0.39	0.51
37	-94.64	0.76	-57.00	0.51	-5.7	0.86	0.71	0.48	0.56
30	-123.6	1.00	-111.2	1.00	-6.6	1.00	1.00	0.47	0.65
32	-109.5	0.88	-95.16	0.85	-6.4	0.96	0.90	0.32	0.52
33	-106.9	0.86	-76.32	0.68	-5.6	0.84	0.79	0.46	0.57
35	-106.3	0.86	-56.69	0.50	-5.6	0.84	0.73	0.48	0.57
36	-93.47	0.75	-64.61	0.58	-5.6	0.84	0.72	0.41	0.52
40	-94.63	0.76	-83.37	0.74	-5.4	0.81	0.77	0.37	0.51
43	-91.60	0.74	-80.04	0.71	-6.6	1.00	0.82	0.47	0.59
45	-92.24	0.74	-80.47	0.72	-5.8	0.87	0.78	0.35	0.50
47	-83.80	0.67	-77.48	0.69	-5.3	0.80	0.72	0.39	0.51
Inhibitor	-117.3	0.94	-74.33	0.66	-6.9	1.04	0.88		

Table 6. Improved results in the combined probability between the prediction model and molecular docking analysis ($Prob_{comb}$) for potential activity against TR. Shown are the compounds considered active (with $Prob_{comb}$ equal to or above 0.50) with values of binding energy, molecular docking probability (Pro_{Dc}), and probability of biological activity ($Prob_{Ac}$).

ID	Moldockscore	$Prob_{Dc}$	Plantscore	$Prob_{Dc}$	Vina	$Prob_{Dc}$	Mean $Prob_{Dc}$	$Prob_{Ac}$	$Prob_{Comb}$
1	-66.32	0.83	-59.59	1	-5.9	0.57	0.80	0.69	0.72
2	-57.65	0.72	-17.88	0.30	-4.9	0.47	0.49	0.67	0.61
3	-70.54	0.88	-23.93	0.40	-5.2	0.50	0.59	0.76	0.70
4	-49.40	0.61	-37.15	0.62	-5.5	0.53	0.59	0.73	0.68
5	-48.52	0.60	-30.76	0.51	-6.1	0.59	0.57	0.75	0.69
6	-51.82	0.64	-12.09	0.20	-5.9	0.57	0.47	0.66	0.59
7	-55.53	0.69	-42.41	0.71	-5.8	0.56	0.65	0.72	0.70
8	-54.15	0.67	-45.39	0.76	-4.5	0.43	0.62	0.79	0.73
9	-38.40	0.48	-9.34	0.15	-5.1	0.49	0.37	0.76	0.62
10	-45.26	0.56	-16.38	0.27	-6.5	0.63	0.49	0.54	0.52
11	-46.73	0.58	-15.93	0.26	-5.8	0.56	0.47	0.64	0.58
12	-62.18	0.77	-11.01	0.18	-4.3	0.41	0.46	0.78	0.66
13	-46.89	0.58	-44.40	0.74	-5.4	0.52	0.61	0.70	0.67
14	-58.57	0.73	-18.65	0.31	-5.3	0.51	0.52	0.76	0.67
15	-45.45	0.56	-10.53	0.17	-5.3	0.51	0.42	0.75	0.63
16	-40.49	0.50	-13.13	0.22	-5	0.04	0.25	0.71	0.55
17	-36.43	0.45	-12.86	0.21	-5.4	0.52	0.39	0.73	0.61
18	-61.88	0.77	-20.42	0.34	-5.9	0.57	0.56	0.65	0.61
19	-61.43	0.76	-37.00	0.62	-5.4	0.52	0.63	0.75	0.71
20	-49.21	0.61	-20.69	0.34	-5	0.04	0.33	0.73	0.59
21	-44.15	0.55	-23.60	0.39	-5.7	0.55	0.50	0.78	0.68
22	-64.33	0.80	-9.64	0.16	-5.9	0.57	0.51	0.75	0.66
23	-40.16	0.50	-16.93	0.28	-5.3	0.51	0.43	0.75	0.63
24	-40.14	0.50	-11.51	0.19	-5.6	0.54	0.41	0.68	0.58

25	-49.90	0.62	-34.74	0.58	-5.1	0.49	0.56	0.78	0.70
26	-49.24	0.61	-32.83	0.55	-5.8	0.56	0.57	0.79	0.71
27	-68.09	0.85	-4.23	0.07	-4.9	0.47	0.46	0.79	0.67
28	-49.14	0.61	-37.18	0.62	-6.1	0.59	0.61	0.78	0.72
29	-32.22	0.40	0.14	0.00	-5.6	0.54	0.31	0.76	0.60
30	-79.81	1	-35.71	0.54	-6.1	0.59	0.73	0.84	0.80
31	-50.03	0.62	-32.6	0.54	-6.1	0.59	0.58	0.80	0.72
32	-59.18	0.74	-20.43	0.34	-6.4	0.62	0.56	0.66	0.62
33	-77.97	0.97	-58.24	0.97	-5.2	0.50	0.81	0.83	0.82
34	-72.31	0.90	-21.00	0.35	-5.8	0.56	0.60	0.78	0.72
35	-65.48	0.82	-24.32	0.40	-5.2	0.50	0.57	0.82	0.73
36	-48.22	0.60	-4.96	0.08	-5.4	0.52	0.40	0.74	0.62
37	-56.29	0.70	-19.97	0.33	-5.7	0.55	0.53	0.69	0.63
39	-62.93	0.78	-12.24	0.20	-5.2	0.50	0.49	0.72	0.64
40	-52.84	0.66	-20.89	0.35	-5.4	0.52	0.51	0.69	0.63
41	-52.96	0.66	-44.53	0.74	-5.2	0.50	0.63	0.73	0.70
42	-62.18	0.77	-37.79	0.63	-6	0.05	0.49	0.78	0.67
43	-66.39	0.83	-28.39	0.47	-10.3	1	0.76	0.82	0.80
44	-53.40	0.66	-13.27	0.22	-6.3	0.61	0.50	0.80	0.69
45	-50.15	0.62	-21.29	0.35	-5.7	0.55	0.51	0.74	0.66
46	-55.91	0.70	-42.10	0.70	-6	0.05	0.48	0.62	0.57
47	-57.64	0.72	-36.96	0.62	-5.2	0.50	0.61	0.72	0.68
Inhibitor	44.76	0.56	-23.02	0.38	-5.6	0.54	0.49		

Table 7. Improved results in the probability of molecular docking ($Prob_{Dc}$) for potential activity against CYP51. Compounds considered active (with values equal to or greater than 0.56 in the $Prob_{Dc}$) are shown.

ID	Moldockscore	$Prob_{Dc}$	Plantscore	$Prob_{Dc}$	Vina	$Prob_{Dc}$	Mean $Prob_{Dc}$
2	-103.5	0.50	-82.78	0.70	-7.3	0.60	0.60
3	-88.14	0.43	-82.24	0.70	-7.2	0.60	0.57
5	-89.88	0.44	-68.66	0.58	-8.1	0.675	0.56
8	-124.1	0.60	-80.00	0.68	-6.2	0.51	0.60
9	-97.84	0.48	-79.03	0.67	-7.4	0.61	0.59
12	-121.2	0.59	-91.58	0.78	-6.6	0.55	0.64
18	-108.3	0.53	-88.22	0.75	-7	0.58	0.62
19	-96.69	0.47	-78.88	0.67	-8.1	0.67	0.60
20	-96.77	0.47	-74.36	0.63	-7.2	0.60	0.56
22	-99.31	0.48	-91.35	0.77	-7.3	0.60	0.62
23	-82.40	0.40	-71.74	0.61	-7.2	0.60	0.53
27	-106.5	0.52	-86.27	0.73	-6.6	0.55	0.60
30	-118.7	0.58	-100.1	0.85	-8.6	0.71	0.71
32	-85.79	0.42	-75.08	0.63	-8.5	0.70	0.58
33	-203.6	1	-117.4	1	-8	0.66	0.88
35	-107.0	0.52	-82.86	0.70	-7.2	0.60	0.61
37	-89.37	0.43	-72.98	0.62	-8.2	0.68	0.58
41	-94.52	0.46	-76.22	0.64	-7.3	0.60	0.57
43	-116.7	0.57	-107.4	0.91	-8.5	0.70	0.73
44	-87.29	0.42	-75.81	0.64	-7.3	0.60	0.56
Inhibitor	-80.67	0.39	-77.65	0.66	-7.5	0.62	0.56

3.5 Interaction analysis

Several neolignans in this study obtained promising bindeing energy values, biological activity, combined analyses, pharmacokinetics and pharmaco-chemical properties, and low toxicity. We chose two compounds (**33** and **43**) that excelled in all of these properties to evaluate the interactions and bonds formed with the amino acids of the active site of the three target enzymes (Figures 4 to 6).

Cruzain

Cruzain is the main cysteine protease in *T. cruzi*, and is essential at all stages of the parasite's development. Overexpression of cruzaine increases the transformation of the parasite into an infectious form and is therefore considered an attractive target in drug planning [34].

Compound **33** formed a hydrophobic interaction with Leu37, a steric interaction with Ala133 and three hydrogen bonds with the amino acids Trg59, Asp60 and Asp158. While compound **43** formed only two hydrogen bonds with the amino acids Ser61 and Asn70, in addition to several hydrophobic interactions with residues Met68, Gly65, Leu67, Gly66 and Ala133. According to Gillmor *et al*. (1997) [35], an irreversible cruzain inhibitor, benzoyl-Tyr-Ala, could interact similarly to neolignans with a cruzain active site. According to the authors, interactions with Met68, Leu67 and Glu205 were important to inhibit the active of the enzyme's active.

TR

Trypanothione reductase, found in the epimastigote and trypomastigote forms of the parasite, is a key enzyme in redox metabolism and is essential for trypanosomes. This enzyme is absent in humans, replaced by glutathione and glutathione reductase, offering a target for selective inhibition.

Compound **33** formed three strong hydrogen bonds with residues Gly14, Ser15 and Glu19. Hydrophobic interactions were also observed with the amino acids Cys53, Tyr111 and Met114. Compound **43** showed more hydrogen bonds compared to compound **33,** forming notable bonds with the amino acids Ser15, Glu19, Ile107, Ser110, Tyr111 and Met114. Only a weaker interaction was observed with the amino acid Ile339. A study by Saravanamuthu *et al*. (2004) [36] showed that a *T. cruzi* inhibitor of TR formed several

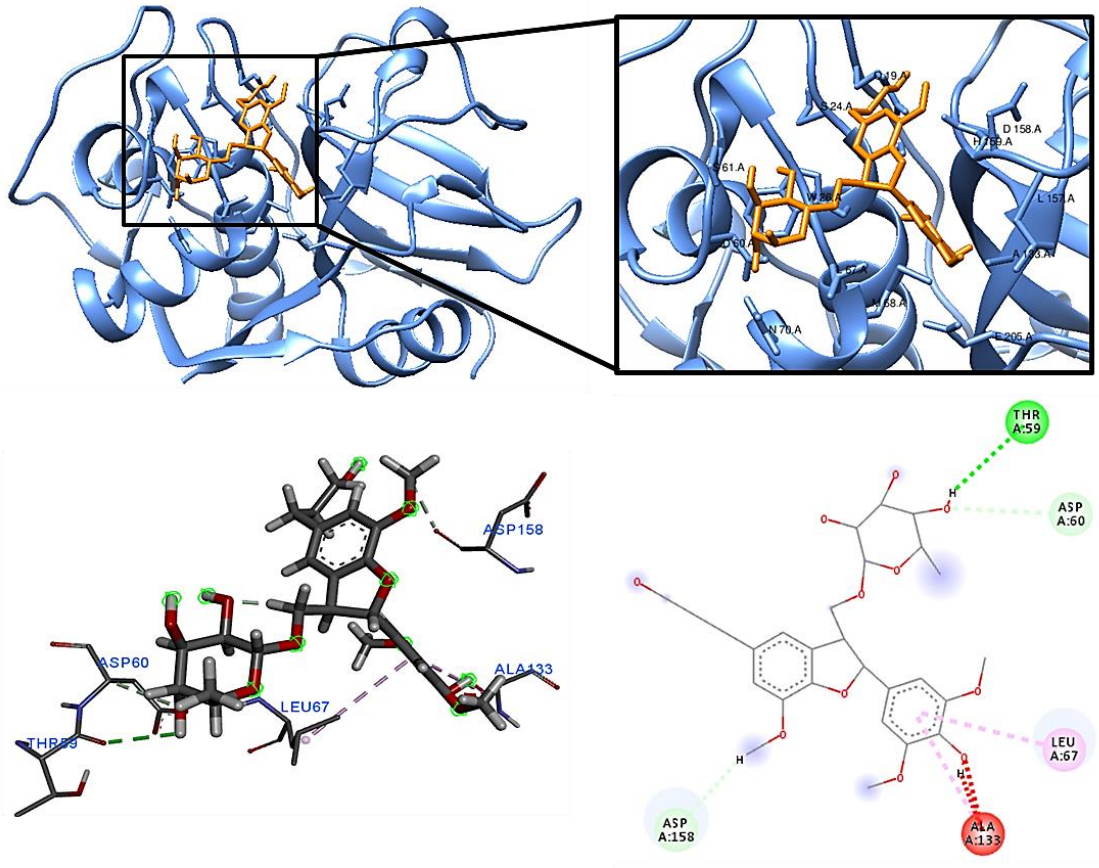
interactions also observed with neolignans and the active site of TR. Notable interactions include those with Glu19, Trp22, Cys53, Ser110, Tyr111, Asp117 and Leu399 stand out.

CYP51

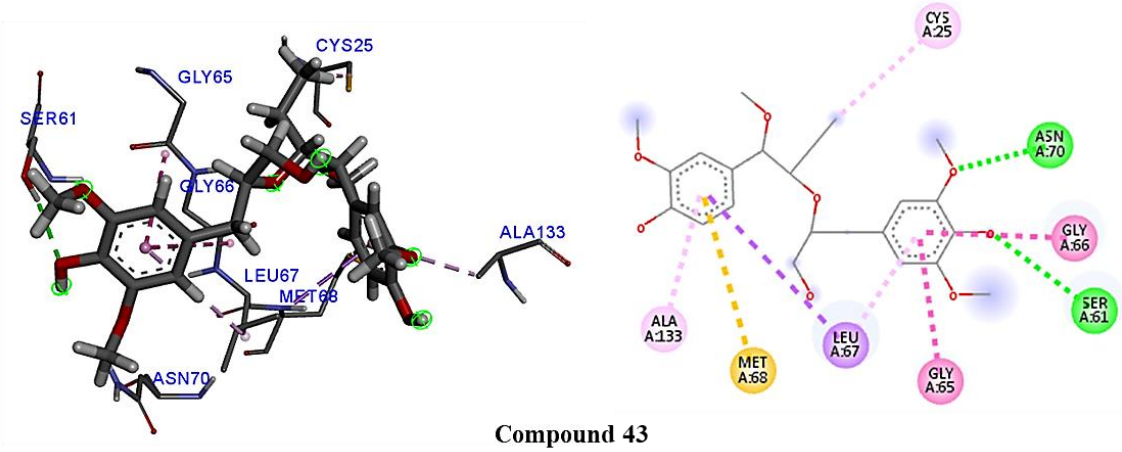
The CYP51 enzyme is involved in a key step in ergosterol biosynthesis, responsible for the oxidized demethylation of intermediate sterols through the heme group. It is essential for the parasite's survival, development and proliferation, which is why it is present in all cellular forms [37].

Compound **33** formed four strong hydrogen bonds with the amino acids Ala287, Ala291, Met358 and Ala414 from the active site of CYP51. It also showed several hydrophobic interactions with the residues Tyr103, Met106, Ala115, Met123, Leu127 and Leu356. While compound **43** showed hydrophobic interactions with residues Phe110, Ala115, Ala287 and Leu356. Two hydrogen bonds with the amino acids Tyr116 and Phe290 and a steric interaction with the amino acid Leu130.

Figure 4. 3D and 2D interactions of neolignans **33** and **43** with the cruzain enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

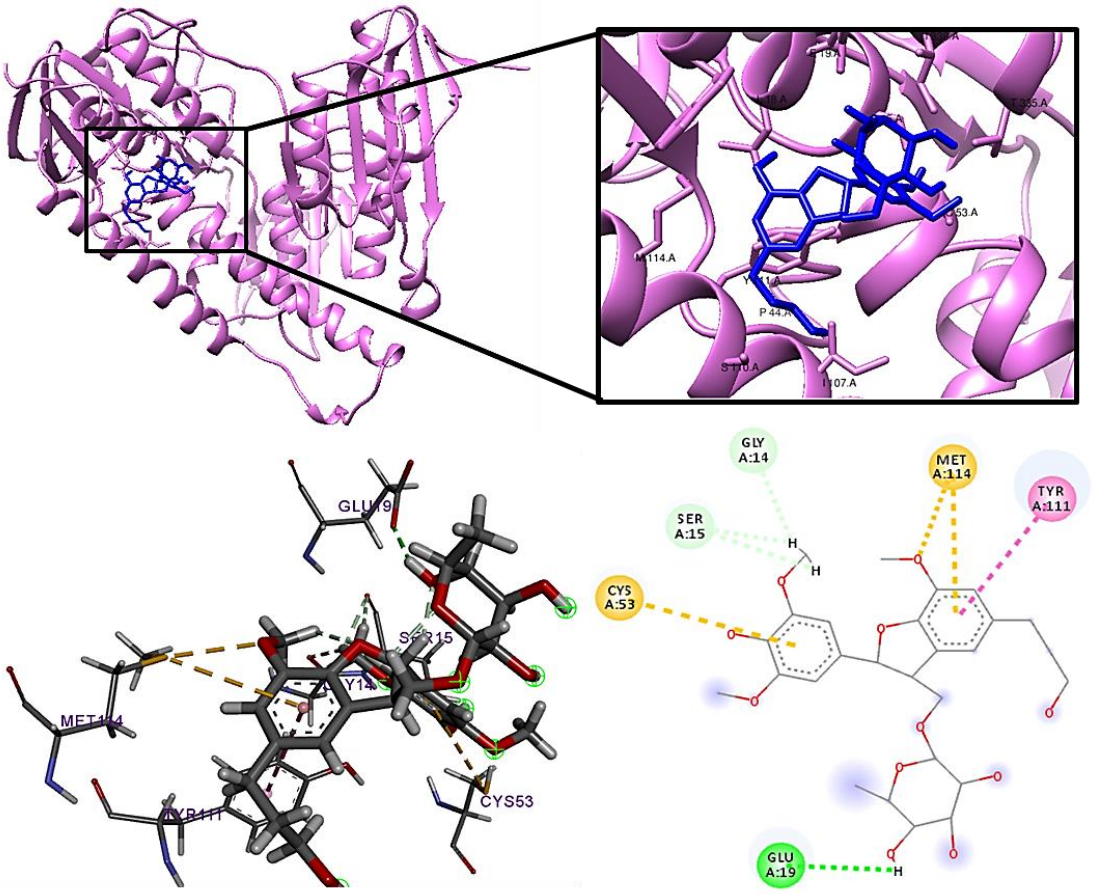


Compound 33

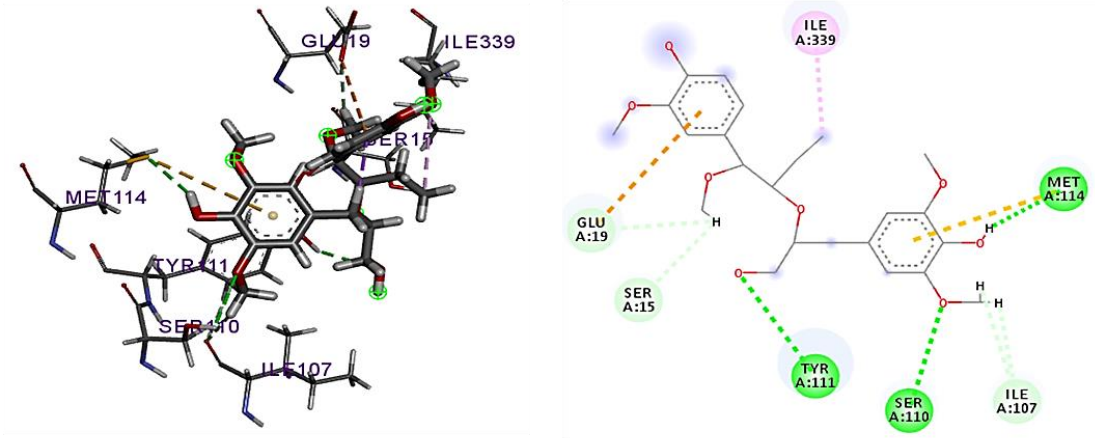


Compound 43

Figure 5. 3D and 2D interactions between neolignans **33** and **43** with the TR enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

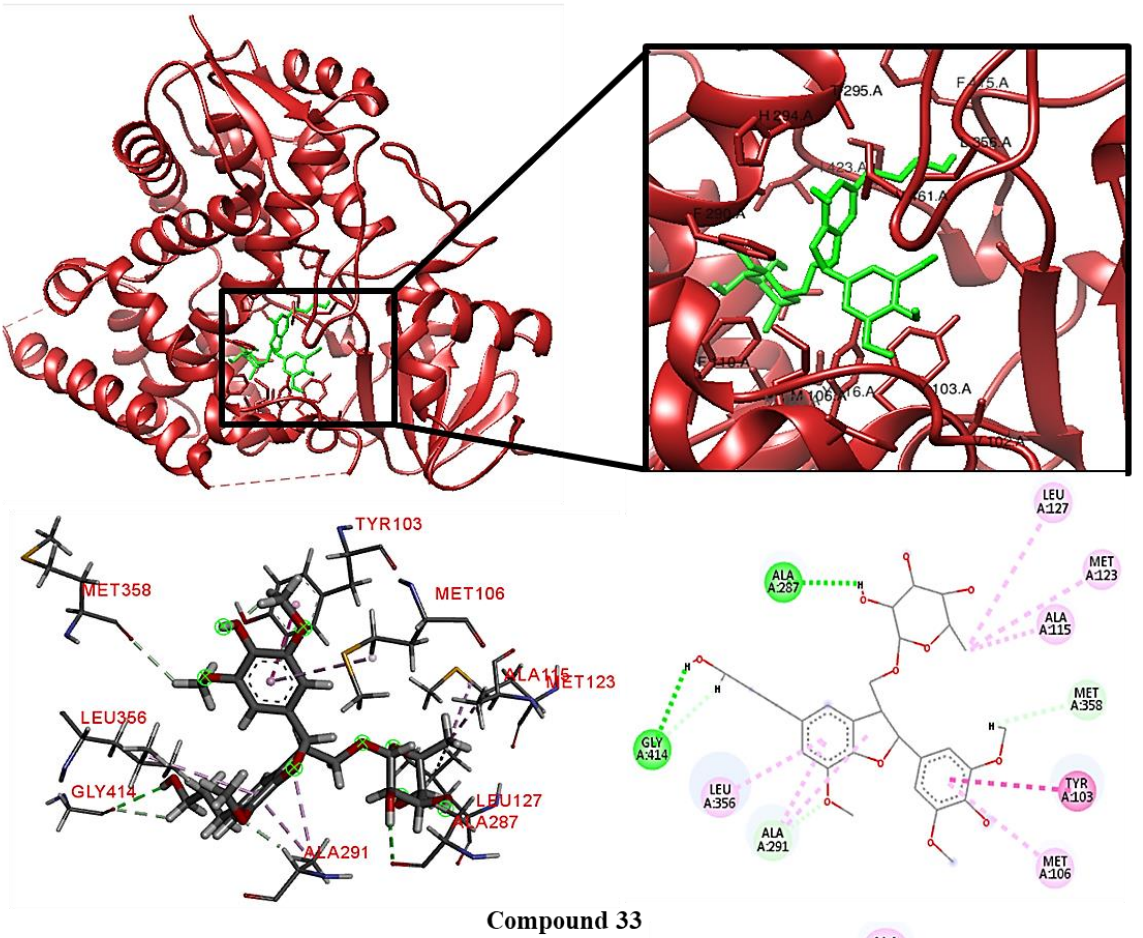


Compound 33

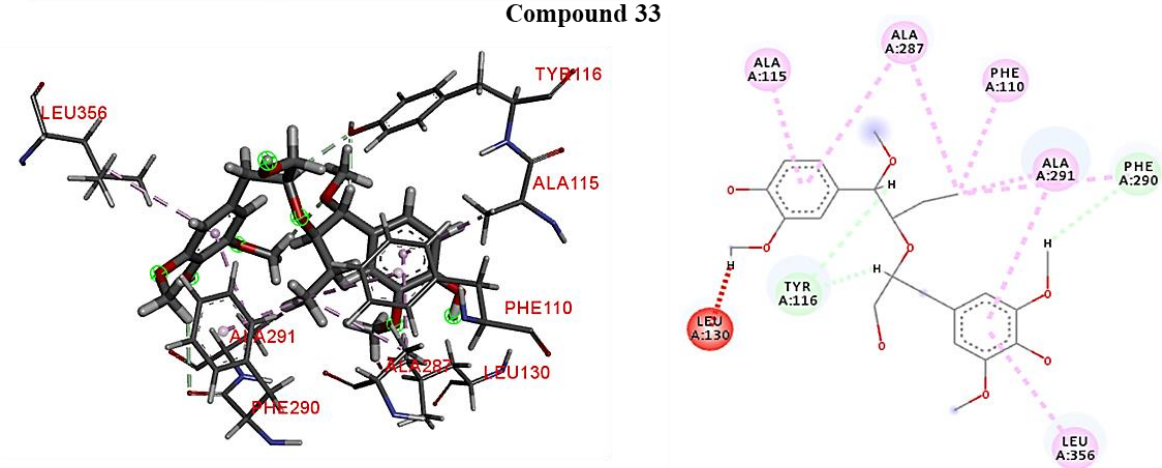


Compound 43

Figure 6. 3D and 2D interactions between neolignan 33 and 43 with the CYP51 enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.



Compound 33



Compound 43

3.6 Molecular dynamics simulations

After the virtual screening and the analysis of the activity potential of several neolignans against important *T. cruzi* enzymes, we conducted molecular dynamics simulations with the two compounds that we were able to isolate (compound **46** and **47**) to assess the flexibility and stability of the enzymes and their interactions in the presence of factors such as solvent, ions, pressure, and temperature. This information is important because it complements the docking results and allows us to evaluate if the compounds remain strongly linked to the studied enzymes in the presence of factors found in the host organism. We chose the TR enzyme for this analysis, because neolignans were more selective for this protein. Then, the RMSD was calculated for the Cα atoms of the complexed enzyme and the structures of each ligand, separately.

The RMSD analysis of the TR enzyme complexed with the crystallographic ligand showed conformations ranging from 0.25 to 0.4 nm in size for 10 ns, with high stability (Figure 7). The same pattern was observed for the enzyme complexed with the neolignans. The stability of this protein is essential to keep compounds bound to the active site.

When we analyzed the flexibility of the ligands, we found that the crystallographic ligand was drastically more unstable than the neolignans during the entire dynamics simulation (Figure 8). Therefore, in the presence of solvents, ions and other factors, neolignans can establish stronger bonds with the active site.

To understand the flexibility of the residues and amino acids that contribute to the conformational changes in the TR enzyme, the mean quadratic fluctuation (RMSF) values were calculated for each amino acid in each enzyme. High RMSF values suggest greater flexibility, while low RMSF values reflect less flexibility. Since amino acids with fluctuations above 0.3 nm contribute to the flexibility of the protein structure, we found that residues at positions 1, 80−90, 460−462, and 486 contribute to conformational changes in the TR enzyme (Figure 9). We also found that none of the amino acids affecting the structural conformations identified in TR were active site components. This helps the neolignans to remain in the active site.

Through graphic programs, it was possible to analyze 2D interactions at different times during the simulation of molecular dynamics (Figure 10). We noticed that most of

the interactions observed in docking, were also observed in the dynamics simulations, that is, even in the presence of solvent and ions. Among the observed interactions, Val58, Ile106, Tyr110 and Met113 are notable.

Figure 7. RMSD values for the Cα atoms of the TR enzyme complexed with neolignans and the Protein Data Bank (PDB) ligand.

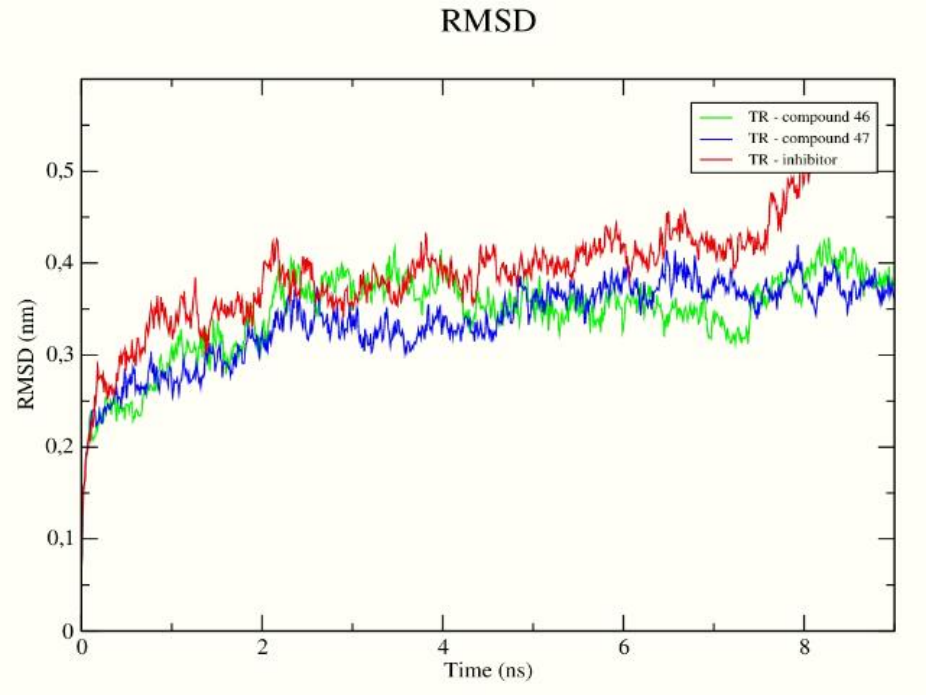


Figure 8. The RMSD values of the Cα atoms of the neolignans and the PDB ligand.

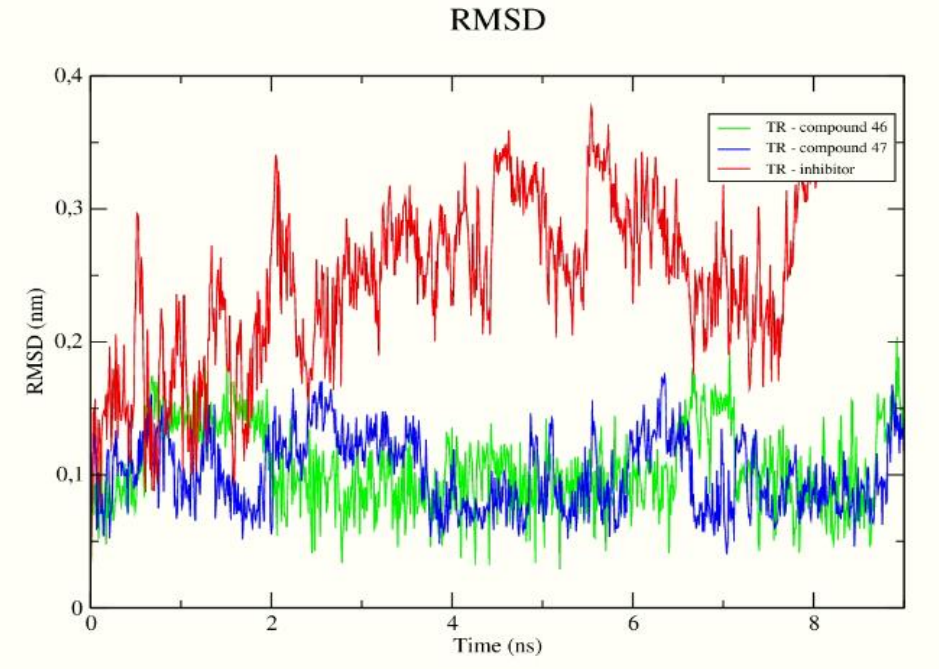


Figure 9. Root-mean-square fluctuation (RMSF) for the Cα atoms of the TR enzyme complexed with the neolignans and the PDB ligand.

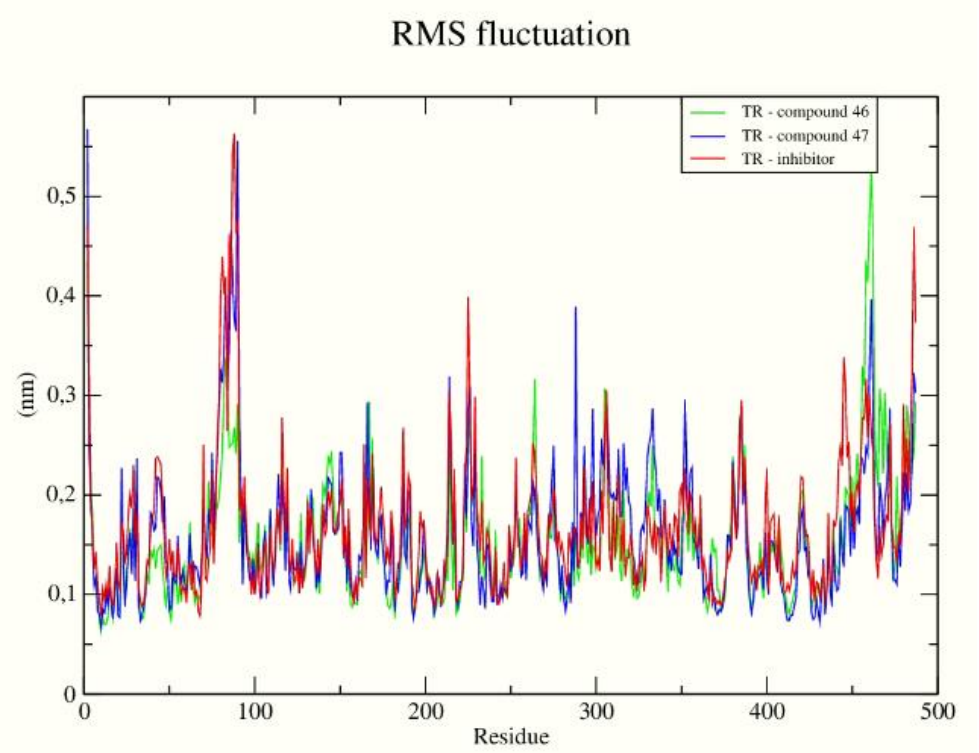
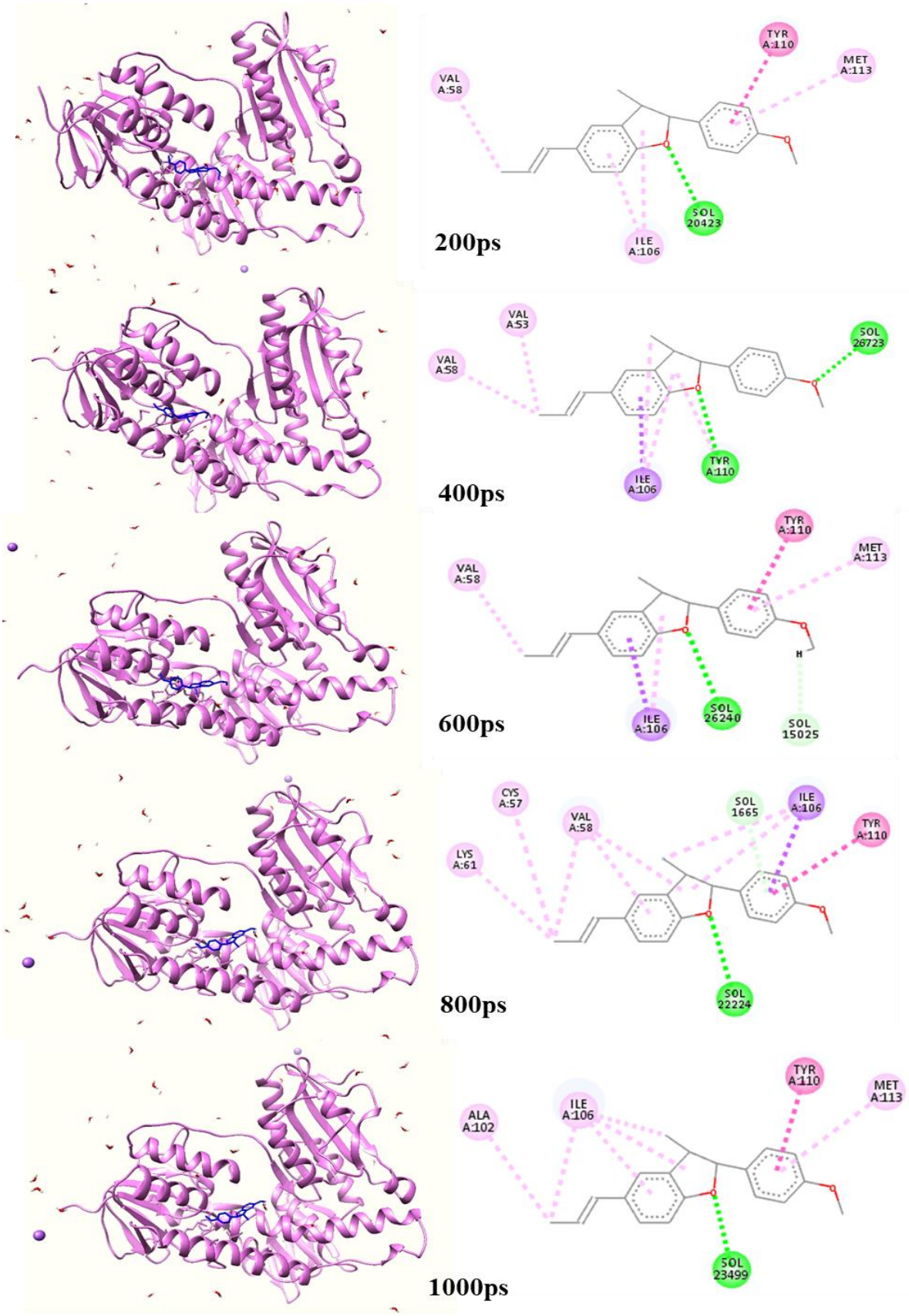


Figure 10. 3D structures and 2D interactions of compound **46** at different times during dynamics simulation.



3.7 Free energy calculations

The MM/PBSA method was used to explore the arrangement of the interactions further and estimate the free energy after the DM simulation. As seen in Table 8, the crystallographic ligand had superior free binding energy to the neolignans that obtained the best results in docking and predicting of biological activity. Although lignans **46** and **47** showed higher values of binding affinity in the docking than the crystallographic ligand, only with the results obtained by the MM/PBSA calculations was it possible to verify that in the presence of a solvent, the inhibitor provided better performance than the neolignans. When analyzing energy contributions, we note that the value of electrostatic and van der Waals energies favored the inhibitor's increased free energy. However, although these results are excellent for the inhibitor, they do not indicate that neolignans did not produce activity, as hydrogen bonds are strong and are not evaluated by the MM/PBSA method.

Table 8. Results of free energy calculations using the MM/PBSA method.

ID	ΔE_{wdw}	$\Delta E_{Electrostatic}$	ΔG_{PB}	ΔG_{SA}	ΔG_{MMPBSA}
46	-99.24 +/- 13.60 kJ/mol	-9.39 +/- 6.847 kJ/mol	43.20 +/- 15.78 kJ/mol	-12.041 +/- 1.43 kJ/mol	-77.46 +/- 14.73 kJ/mol
47	-108.26 +/- 14.39 kJ/mol	-4.73 +/- 3.46 kJ/mol	42.03 +/- 13.70 kJ/mol	-13.23 +/- 1.61 kJ/mol	-84.20 +/- 11.98 kJ/mol
Inhibitor	-151.46 +/- 29.15 kJ/mol	-139.33 +/- 59.85 kJ/mol	112.97 +/- 89.26 kJ/mol	-14.48 +/- 2.92 kJ/mol	-192.31 +/- 37.04 kJ/mol

3.8 Activity of two neolignans against epimastigotes and trypomastigotes *T. cruzi*

We isolated two neolignans that obtained excellent results during the virtual screening process. The compounds were evaluated for the potential to inhibit the growth of epimastigotes and trypomastigotes. The results showed that the neolignans (2R, 3R) -2,3Di-hydro-2- (4-methoxyphenyl) -3-methyl-5- (E) -propenylbenzofuran (**46**) and ottomentosa (**47**) were able to potentially inhibit both cellular forms of *T. cruzi*, with inhibitory activity superior to benznidazole (Table 9). The best results were against trypomastigote forms, which correspond to the infectious phase of CD. The ottomentosa neolignan (**47**) proved to be lightly more potent than the neolignan (2R, 3R) -2,3Di-hydro-2- (4-methoxyphenyl) -3-methyl-5- (E) -propenylbenzofuran (**46**) . Therefore, both substances acted as therapeutic trypanomicidal agents, as they could kill parasites without affecting renal cells.

Table 9. Effect on *Trypanosoma cruzi* and renal cells.

ID	Epimastigote IC_{50} (µM)	Trypomastigote IC_{50} (µM)	Cytotoxicity (cells LLC-MK2) (µM)
46	9.64 ± 1.75	4.88 ± 0.61	> 400
47	8.72 ± 1.14	2.73 ± 0.58	> 400
Benznidazol	16.58 ± 2.9	257.5 ± 14.32	> 400

Some previous studies have demonstrated the trypanomicidal activity of neolignans, but this is the first reported describing the activity of several specific neolignans. A study by Cabral *et al.* (2010) [38] showed that two neolignans, licarin A and burchellin, could inhibit the growth of the epimastigote by 45% and 20%, respectively. The authors also found that lycarin A and burchellin could induce trypomastigote death with IC_{50}/24h of 960 µM and 520 µM, respectively. Pelizzaro-Rocha *et al*. (2011) [39] demonstrated that the neolignan eupomatenoid-5 exhibited activity against trypomastigotes, the infective form of *T. cruzi* (EC_{50} 40.5 µM), leading to ultrastructural alteration and lipoperoxidation in the cell membrane. In addition, they reported that the trypanocidal action of eupomatenoid-5 might be associated with mitochondrial dysfunction and oxidative damage, which can trigger destructive effects on the biological molecules of *T. cruzi*, leading to the death of the parasite. Ferreira *et al*. (2019) used a semi-synthetic library of 23 derivatives of the neolignan dehydrodieugenol B that was prepared to explore synthetically accessible activity structure (SAR) relationships against *T. cruzi*. Five compounds demonstrated activity against trypomastigotes (IC_{50} values from 8 to 64 µM) and eight showed activity against intracellular amastigotes (IC_{50} values from 7 to 16 µM) [40].

4. CONCLUSIONS

We used a comprehensive computational studies approach to investigate the potential of neolignans in the treatment of CD, which made possible the isolation and experimental testing of natural products against cellular forms of *T. cruzi*. The predictive models generated from essential enzymes of the parasite obtained excellent performance results, with an accuracy greater than 75%, and selected a neolignan with a 50%

probability of active potential for the cruzain enzyme. For the TR enzyme, an accuracy of 85% was achieved, and the model selected all neolignans, with activity probabilities between 54% and 84%. Therefore, neolignans were considered selective against the TR enzyme.

For a structure-based investigation, a consensus docking analysis was conducted to ensure the reliability of the RF model and to reduce the number of false positives. Among the 47 neolignans analyzed by molecular docking, two compounds were considered potentially active against cruzain, 31 neolignans active against TR, and 18 against CYP51. These results indicated that neolignans, in general, are more likely to activate TR and CYP51 proteins, and are not selective for the cruzain enzyme.

A structure and ligand-based combined analysis, employed to increase the predictive power was able to identify potentially active molecules, using RF models and molecular docking, resulting in the identification of 22 compounds with a probability of activity ranging from 50%−65% for cruzain and 46 neolignans potentially active against the TR enzyme with a probability ranging from 52−82%. We found that the combined analysis expanded the selection of active compounds for cruzain relative to the RF model and molecular docking. We also found that neolignans were more selective for the TR enzyme.

MD simulations revealed that neolignan-complexed RT was stable under several conditions, including solvent, ions, temperature, and pressure, with only small variations observed for some complexed compounds. Therefore, the binding affinity between proteins and ligands is unlikely to be affected by environmental changes. In addition, none of the amino acids responsible for the enzymatic conformational changes were at the active site, allowing the active site to remain stable. In addition, through calculations of free energy using the MM/PBSA method, we found that although the crystallographic ligand presented a higher energy bond value in the presence of factors, it is notable that strong hydrogen bonds also favor the permanence of the neolignans at the active site of TR.

Two neolignans with excellent ADMET profiles demonstrated to be potentially active inhibitors against the enzymes cruzain and TR enzymes by virtual screening, were isolated from *Krameria tomentosa* and subjected to in vitro tests. The results showed that

the two neolignans (**46** and **47**) could potentially inhibit *T. cruzi*, at concentrations of 9.64 and 8.72 µM for the epimastigote forms and 4.88 and 2.73 µM for the trypomastigote forms, respectively. Therefore, the compounds compounds (2R, 3R)-2,3-Di-hydro-2-(4-methoxyphenyl)-3-methyl-5-(E)-propenylbenzofuran (**46**) and ottomentosa (**47**) proved to be excellent inhibitors of growth for the epimastigote and promastigote stages of the parasite.

We also concluded that the neolignans investigated in this study that were considered active against the enzymes cruzain and CYP51 could be potent inhibitors of these enzymes in amastigotes, since these enzymes are present in all cellular forms of *T. cruzi*. Therefore, the selected neolignans should be investigated in further studies to treat the chronic phase of DC.

REFERENCES

1. Lidani KCF, Andrade FA, Bavia L, Damasceno FS, Beltrame MH, Messias-Reason IJ, et al. Chagas disease: From discovery to a worldwide health problem. J Phys Oceanogr. 2019;49(6):1–13.

2. Gasparim AZ, Fontes CER, Rossoni DF, Toledo MJ de O. Epidemiological and clinical profile of patients with chagas disease in the central-north area of Paraná, Southern Brazil. Rev Soc Bras Med Trop. 2018;51(2):225–30.

3. Villalta F, Rachakonda G. Advances in preclinical approaches to Chagas disease drug discovery. Expert Opin Drug Discov [Internet]. 2019;14(11):1161–74. Available from: https://doi.org/10.1080/17460441.2019.1652593

4. Ferri G, Edreira MM. All Roads Lead to Cytosol: Trypanosoma cruzi Multi-Strategic Approach to Invasion. Front Cell Infect Microbiol. 2021;11(March):1–17.

5. Meymandi S, Hernandez S, Park S, Sanchez DR, Forsyth C. Treatment of Chagas Disease in the United States. Curr Treat Options Infect Dis. 2018;10(3):373–88.

6. Moura M De, Houten B Van. Unveiling Benznidazole's Mechanism of Action Through Overexpression of DNA Repair Proteins inTrypanosoma cruzi. Environ Mol Mutagen [Internet]. 2010;405(April):391–405. Available from: http://dx.doi.org/10.1002/em.20575

7. Vinuesa T, Herráez R, Oliver L, Elizondo E, Acarregui A, Esquisabel A, et al. Benznidazole nanoformulates: A chance to improve therapeutics for Chagas disease. Am J Trop Med Hyg. 2017;97(5):1469–76.

8. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al.

ChEMBL: A large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012;40(D1):1100–7.

9. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: An update. Nucleic Acids Res. 2014;42(D1):1083–90.

10. Fourches D, Muratov E, Tropsha A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. J Chem Inf Model. 2010;50(7):1189–204.

11. dos Santos Maia M, Silva JPR, de Lima Nunes TA, de Sousa JMS, Rodrigues GCS, Monteiro AFM, et al. Virtual screening and the in vitro assessment of the antileishmanial activity of lignans. Molecules. 2020;25(10):1–33.

12. Dos Santos Maia M, Rodrigues GCS, De Sousa NF, Scotti MT, Scotti L, Mendonça-Junior FJB. Identification of New Targets and the Virtual Screening of Lignans against Alzheimer's Disease. Oxid Med Cell Longev. 2020;2020.

13. S. Maia M, de Sousa NF, Rodrigues GCS, Monteiro AFM, Scotti MT, Scotti L. Lignans and Neolignans Anti-tuberculosis Identified by QSAR and Molecular Modeling. Comb Chem High Throughput Screen. 2020;23(6):504–16.

14. Mauri A, Consonni V, Pavan M, Todeschini R. DRAGON software: An easy approach to molecular descriptor calculations. Match. 2006;56(2):237–48.

15. Gramatica P. Principles of QSAR Modeling. Int J Quant Struct Relationships. 2020;5(3):1–37.

16. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The protein data bank: A computer-based archival file for macromolecular structures. Arch Biochem Biophys. 1978;185(2):584–91.

17. Thomsen R, Christensen MH. MolDock: A new technique for high-accuracy molecular docking. J Med Chem. 2006;49(11):3315–21.

18. Trott, O.; Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–461.

19. Sargis Dallakyan; Arthur J. Olson. Small-Molecule Library Screening by Docking with PyRx. Chem Biol. 2014;243–50.

20. Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep [Internet]. 2017;7(October 2016):1–13. Available from: http://dx.doi.org/10.1038/srep42717

21. Sander T, Freyss J, Von Korff M, Reich JR, Rufener C. OSIRIS, an entirely in-house developed drug discovery informatics system. J Chem Inf Model. 2009;49(2):232–46.

22. Achenbach H, Utz W, Usubillaga A, Rodriguez HA. Lignans From K R a M E R I a Ixina *. 1991;30(11):3753–7.

23. Madeiro SAL, De Lucena HFS, Siqueira CD, Duarte MC, Braz-Filho R, Filho JMB, et al. New neolignans from krameria tomentosa A. St.-Hil. J Braz Chem Soc. 2012;23(11):2021–6.

24. Sawasdee K, Chaowasku T, Likhitwitayawuid K. New neolignans and a phenylpropanoid glycoside from twigs of miliusa mollis. Molecules. 2010;15(2):639–48.

25. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25.

26. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1–3):43–56.

27. Bondi. A. Van der Waals Volumes and Radii. J Phys. 1964;68:441–51.

28. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.

29. Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 2014;54(7):1951–62.

30. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001;98(18):10037–41.

31. Aparicio IM, Scharfstein J, Lima APCA. A new cruzipain-mediated pathway of human cell invasion by Trypanosoma cruzi requires trypomastigote membranes. Infect Immun. 2004;72(10):5892–902.

32. Lima DB, Sousa PL, Torres AFC, Rodrigues KA da F, Mello CP, Menezes RRPPB de, et al. Antiparasitic effect of Dinoponera quadriceps giant ant venom. Toxicon. 2016;120:128–32.

33. Mandal S, Moudgil M, Mandal SK. Rational drug design. Eur J Pharmacol [Internet]. 2009;625(1–3):90–100. Available from: http://dx.doi.org/10.1016/j.ejphar.2009.06.065

34. Gillmor SA, Craik CS, Fletterick RJ. Structural determinants of specificity in the cysteine protease cruzain. Protein Sci. 1997;6(8):1603–11.

35. Durrant JD, Keränen H, Wilson BA, McCammon JA. Computational identification of uncharacterized cruzain binding sites. PLoS Negl Trop Dis. 2010;4(5).

36. Saravanamuthu A, Vickers TJ, Bond CS, Peterson MR, Hunter WN, Fairlamb AH. Two interacting binding sites for quinacrine derivatives in the active site of trypanothione reductase: A template for drug design. J Biol Chem [Internet]. 2004;279(28):29493–500. Available from: http://dx.doi.org/10.1074/jbc.M403187200

37. França RRF, De Carvalho AS, Branco FSC, Pinto AC, Boechat N. Potent inhibitors of the enzyme sterol 14α-demethylase against Trypanosoma cruzi. Rev Virtual Quim. 2014;6(5):1483–516.

38. Cabral MMO, Barbosa-Filho JM, Maia GLA, Chaves MCO, Braga M V., De Souza W, et al. Neolignans from plants in northeastern Brazil (Lauraceae) with activity against Trypanosoma cruzi. Exp Parasitol [Internet]. 2010;124(3):319–24. Available from: http://dx.doi.org/10.1016/j.exppara.2009.11.007

39. Pelizzaro-Rocha KJ, Veiga-Santos P, Lazarin-Bidóia D, Ueda-Nakamura T, Dias Filho BP, Ximenes VF, et al. Trypanocidal action of eupomatenoid-5 is related to mitochondrion dysfunction and oxidative damage in Trypanosoma cruzi. Microbes Infect. 2011;13(12–13):1018–24.

40. Ferreira DD, Sousa FS, Costa-Silva TA, Reimão JQ, Torrecilhas AC, Johns DM, et al. Dehydrodieugenol B derivatives as antiparasitic agents: Synthesis and biological activity against Trypanosoma cruzi. Eur J Med Chem. 2019;176:162–74.

Capítulo 4

A Esquistossomose é uma doença negligenciada que afeta milhões de pessoas em vários países do mundo. Embora os tratamentos quimioterápicos sejam eficientes, um dos maiores desafios é a resitência ao medicamento. Por isso, faz-se necessário a busca por novos agentes terapêuticos para o tratamento da Esquistossomose.

Neste capítulo, foi realizada uma análise de perfil farmacocinético com lignanas obtidas do banco de dados ChEMBL. Os compostos com melhores perfis foram submetidos ao docking consenso. Em seguida, o melhores resultados foram submetidos à previsão de atividade biológica a partir de dois modelos QSAR contra *Schistossoma mansoni* em diferentes programas. Uma análise consenso foi realizada para selecionar os compostos mais potentes.

Análises de dinâmica molecular e cálculos de energia livre foram realizadas para investigar a flexibilidade estrutural do complexos e a afinidade de ligação na presença de solvente.

In silico identification of lignans with potential anti-schistosomiasis activity

Mayara dos Santos Maia [1], Renata Priscila Costa Barros [1], Gabriela Cristina Soares Rodrigues [1], Luciana Scotti [1] and Marcus Tullius Scotti [1]*

[1] Laboratory of Cheminformatics, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, João Pessoa-PB, Brazil; mayarasmaia@hotmail.com, renatapcbarros@gmail.com, gaby.ecologia@gnail.com, luciana.scotti@gmail.com and mtscotti@gmail.com.

*Corresponding author: mtscotti@gmail.com.

Artigo será submetido na revista: Saudi Pharmaceutical Journal

Fator de impacto: 4.33

ABSTRACT: Schistosomiasis is a neglected disease that affects low-income populations and can lead to death due to the complications during chronic infection. Although nowadays there are two drugs available for treatment of schistosomiasis, Praquizantel (PZQ) and Oxamniquine (OXA), drug therapy is ineffective for juvenile worms, which leads to resurgence of symptoms. In addition, there are several cases of drug resistance described in the literature. Thus, to discover potent inhibitors against Sulfotransferase (SULT), an important target for *Schistosoma* species, we aimed to *in sili*co identification of lignans with high therapeutic potential for the treatment of schistosomiasis. To achieve this goal, we have used both ligand-based and structure-based approaches. Our results have demonstrated that 28 compounds showed a good ADMET (Absorption, distribution, metabolism, excretion and toxicity) profile. In the consensus docking, 17 molecules obtained predicted values of binding energy higher than OXA. Seven of these molecules were predicted to be active against *Schistosoma mansoni* by consensus model developed by Random Forest and MuDRA algorithms. Molecular dynamics simulations and free energy calculations of selected best compounds indicated that in the presence of solvent, OXA presents higher binding energy

values compared to lignans. However, in cases of drug resistance, lignans 57-60 may serve as an alternative for the treatment of Schistosomiasis.

Keywords: Schistosomiasis, Lignans, ADMET, Consensus docking, QSAR and Molecular dynamics simulations.

1. INTRODUCTION

Schistosomiasis is a parasitic disease caused by flatworms of the Trematoda class. Currently, there are six major Schistosoma species that infect humans: *Schistosoma mansoni*, *S. haematobium*, *S. intercalatum*, *S. guineensis*, *S. mekongi*, and *S. japonicum*; two latter species being recognized by zoonosis (Leger & Webster, 2017; Porcino et al., 2019; Mäder et al., 2018). Adult worms inhabit the blood vessels of vertebrate hosts, but their life cycle begins with asexual multiplication or development within an invertebrate intermediate host - a snail (*Biomphalaria* spp., *Bulinus* spp. and *Oncomelania* spp.). Transmission in humans occurs by contact with freshwater contaminated by cercariae, which corresponds to the infectious phase of schistosomes that are released by the snail. Acute infection has symptoms such as abdominal pain, diarrhoea (with/without blood), fatigue, malaise, fever, and in, case of S. haematobium infection, hematuria (McManus et al., 2018). Chronic infection can cause anemia, stunted growth, impaired cognition, decreased physical fitness, intestinal fibrosis, hepatosplenomegaly, neurological complications, and, ultimately, death (Zoni et al., 2016).

S. mansoni is endemic in sub-Saharan Africa, Brazil, Caribbean islands, Puerto Rico, Suriname, and Venezuela, occupying the largest risk area. It is estimated that more than 700 million people are in risk areas (da Silva et al., 2017).

Several vaccine candidates against Schistosomiasis in humans are still in the testing phase (Merrifield et al., 2016). The current treatment is done with two available chemotherapeutic agents, Praquizantel (PZQ) and Oxamniquine (OXA). PZQ is the drug most used in the treatment of schistosomiasis, reaching cure rates of 60 to 70%, but it is ineffective in immature worms and juveniles, which contributes to treatment failures and resurgence of symptoms. OXA represents an alternative, but is only active against *S. Mansoni* (Mäder et al., 2018). OXA showed a high cure rate and no serious side effects, although stomach discomfort, dizziness and headaches may occur. People with epilepsy

should not take this medicine. The treatment fails because of low activity in immature worms and drug resistance (da Silva et al., 2017).

Exploring the mechanism of action of OXA, researchers have discovered that this drug is sulfonated by an endogenous Sulfotransferase (SULT) (da Silva et al., 2017). The SULT of Schistossoma spp. adopts an α/β fold characteristic to the 3'-Phosphoadenosine-5'-phosphosulfate (PAPS) dependent SULT family of enzymes with a parallel β-sheet of five filaments intercalated between two groups of five and seven α-helices and its activity consists of the nucleophilic attack of the hydroxyl group of the deprotonated substrate in the sulfur of PAPS (Taylor et al., 2017).

Knowing the importance of SULT for the schistosome parasite, its selectivity for the genus, and the limitations of the medicines available for the treatment, given study aimed to perform computer-assisted discovery of lignans, natural products that may contribute as alternatives of treatment for Schistosomiasis.

2. MATERIAL AND METHODS

2.1 Data collection and curation

Chemical compounds with known activity against S. mansoni (pIC_{50}) were obtained from the ChEMBL database. All the compounds with $pIC_{50}>6$ were considered as active and all other as inactive, resulting in 129 actives and 180 inactives. In addition, we obtained 155 lignans from CHEMBL (https://www.ebi.ac.uk/chembl/) to use for virtual screening. All the compounds were carefully curated for both chemical and biological data according to the workflows described by Fourches et al (Fourches et al., 2016; Fourches et al., 2015; Fourches et al., 2010). Duplicate search was done using HiT QSAR software (Tomasz Puzyn , Jerzy Leszczynski, 2009). Three-dimensional structures were generated by ChemaxonStandardiser v.18.17.0, (www.chemaxon.org).

2.2 Predicting ADMET properties

The ADME parameters were calculated using the SwissADME free access web tool (http://www.swissadme.ch). Toxicity prediction was performed in OSIRIS Property Explorer

(https://www.organic-chemistry.org/prog/peo/) (Mandal et al., 2009). For absorption, factors including membrane permeability, intestinal absorption and P-glycoprotein substrate or inhibitor were considered. Thus, we investigated compounds with the following properties: molecular weight between 150 and 500 g/mol, topological polar surface area (TPSA) between 20 and 120Å, molecule with up to nine rotational bonds (Daina et al., 2017), do not exceed more than two breaches of the Lipinski and logP consensus not greater than 4.15 and which is not substrate for the permeability glycoprotein (gp-P) enzyme. The distribution was assessed by factors including the blood-brain barrier (BBB) and Central Nervous System (CNS) permeability. Metabolism is predicted based on CYP models for substrate or inhibition (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4). Drug toxicity was predicted based on the following parameters evaluated as: mutagenicity, tumorogenicity, reproductive effect, and irritability.

2.3 Alignment of protein sequences

The sequences of the enzymes Sulfotransferases of the species of *Schistosoma* genus were obtained by the database GenBank (https://www.ncbi.nlm.nih.gov/) (Benson et al., 2013). The crystallographic structures were obtained by the Protein Data Bank - PDB (https://www.rcsb.org/pdb/home/home.do) (Bernstein et al., 1978) with the codes PDB ID 4MUB for *S. mansoni*, PDB ID 5TIX for *S. haematobium*, and PDB ID 5TIZ for *S. japonicum*. Then, the global alignment was performed through the Clustal Omega web tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) that aligns all the protein sequences entered by the user. Terminal regions of sequences that are not shared were omitted.

2.4 Consensus docking

The docking was performed using four different packages, i.e., Molegro Virtual Docker (MVD) (Thomsen & Christensen, 2006), GOLD 5.6.2 (Verdonk et al., 2003), AutoDock Vina (Vina) (Sargis Dallakyan; Arthur J. Olson, 2014), and AutoDock 4.2.6. (AD4), with default parameters (Morris et al., 2010). Sulfotransferase (SULT) enzyme of *S. mansoni* (PDB ID 4MUB) was selected for docking. Initially, all water molecules were removed from the crystalline structure. Root mean square deviation (RMSD) ≤ was used as a criterion of success. The consensus strategy consisted in selecting compounds with higher

binding affinity prediction values than crystallographic binders in at least three docking programs. This type of consensus methodology contributes to increase the reliability of docking and increase the number of true positive compounds.

2.5 QSAR modeling

The Knime 3.6.2 software (Knime 3.4.0, Copyright Miner, from Konstanz Information, 2003-2017, www.knime.org) was used to perform Quantitative structure–activity relationship (QSAR) modeling. Given success of our previous studies (Lorenzo et al., 2015; Melo-Filho et al., 2016), we decided to perform 3-D QSAR analysis. To do so, all the studied compounds with curated chemical structure were saved in SDF format and imported into VolSurf + v.1.0.7 software (Cruciani et al., 2000). There, the GRID force field was used. Random forest implementation described by us earlier1 was used for building the models. Applicability domain was estimated accordingly to the procedures described elsewhere (Dos Santos Maia M; de Sousa NF; Rodrigues GCS; Monteiro AFM; Scotti MT; Scotti L., 2020). Five-fold external cross-validation was utilized for estimation of predictive power of developed models. In addition to RF, we also used simpler and more illustrative MuDRA (Alves et al., 2018) approach for building the models.

2.6 Molecular dynamics (MD)

The molecular dynamics simulations were performed to estimate the flexibility of the interactions between proteins and ligands using GROMACS 5.0 software (Abraham et al., 2015; Berendsen et al., 1995). For this, the compounds 57, 58, 59 and 60 complexed to SULT were selected because of the excellent results in docking and prediction of biological activity using the Random Forest and MuDRA models. The topology of the ligands was prepared using the PRODRG topology generator (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg/submit.html) (Schüttelkopf & Van Aalten, 2004) applying the force field GROMOS. The topology of the protein was also prepared using the GROMOS force field. The molecular dynamics simulation was performed using the SPC water model of point load extended in a cubic box (Bondi., 1964). The system was neutralized by the addition of ions (Cl- and/or Na+) and minimized to remove bad contacts between complex molecules and solvent. The system was also balanced in 300K using the V-rescale algorithm at 100ps

represented by NVT (constant number of particles, volume and temperature), followed by equilibrium at 1 atm of pressure using the Parrinello-Rahman algorithm as NPT (constant pressure numerical particles and temperature) to 100 ps. The DM simulations were performed in 5,000,000 cycles at 10 ns. To determine the flexibility of the structure and if the complex is stable near the experimental framework, the mean square root displacement (RMSD) of all α-carbon atoms was calculated in relation to the starting structures. Root Mean Square Fluctuation (RMSF) were also analyzed to understand the role of residues near the receptor binding site. RMSD and RMSF plots were generated in Grace Software (http://plasma-gate.weizmann.ac.il/Grace/), protein and ligands were visualized at UCSF Chimera (Pettersen et al., 2004). The Surface Area Method of the Molecular Mechanism - Poisson Boltzmann Surface Area (MM/PBSA) was used to calculate the binding-free energy of the protein-binding complex in the study of the molecular behavior of the SULT enzyme and its respective ligands (Kumari et al., 2014).

3. RESULTS AND DISCUSSION

3.1 Predicting ADMET properties

Twenty eight chemicals with selected after prediction of their ADMET profile are presented in Table 1. According to the rule of Lipinski (Lipinski et al., 2012), molecules with molecular weight less than 500 Da, calculated LogP (CLogP) less than 5, total polar surface area (TPSA) ≤140 Å and rotatable ≤10 bonds exhibit excellent absorption and bioavailability. Regarding metabolism and toxicity, we chose compounds that presented high gastrointestinal absorption, did not present encephalic permeability, and are substrates for at most two CYP enzymes. Metabolism can have significant effects on the efficacy of the drug, acting on its bioavailability, half-life and generation of harmful metabolites or without therapeutic potential.

Toxicity was also evaluated, and the 28 selected compounds (see Table 1) were predicted not to have negative effect on the reproductive system and irritability, little or no mutagenicity, tumorigenesis, negative effect on the reproductive system and irritability.

Table 1. Lignans with better ADMET profile.

ID	Log P	GI[a]	BBB[b]	P-gp[c]	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4
1	1.24	High	No	Yes	No	No	No	No	No
8	1.3	High	No	Yes	No	No	No	No	No
20	2.38	High	No	Yes	No	No	No	Yes	No
30	2.01	High	No	Yes	No	No	No	Yes	No
31	2.54	High	No	Yes	No	No	No	Yes	No
32	2.52	High	No	Yes	No	No	No	Yes	No
44	2.63	High	No	Yes	No	No	No	Yes	No
57	2	High	No	Yes	No	No	No	Yes	No
58	2.68	High	No	Yes	No	No	No	Yes	No
59	2.88	High	No	Yes	No	No	No	Yes	No
60	2	High	No	Yes	No	No	No	Yes	No
61	2.67	High	No	No	No	No	No	Yes	No
76	3.03	High	No	No	No	No	No	Yes	No
77	2.8	High	No	No	No	No	No	Yes	No
83	2.32	High	No	Yes	No	No	No	Yes	No
84	2.68	High	No	Yes	No	No	No	Yes	No
86	1.5	High	No	Yes	No	No	No	No	No
87	2.18	High	No	Yes	No	No	No	Yes	No
88	1.59	High	No	Yes	No	No	No	No	No
90	1.26	High	No	Yes	No	No	No	No	No
91	2.35	High	No	Yes	No	No	No	Yes	No
121	1.69	High	No	Yes	No	No	No	No	No
124	2.57	High	No	Yes	No	No	Yes	No	No
131	2.51	High	No	No	No	No	No	Yes	Yes
132	1.59	High	No	Yes	No	No	No	No	No
139	1.6	High	No	Yes	No	No	No	No	No
140	1.55	High	No	Yes	No	No	No	No	No
151	2.1	High	No	Yes	No	No	No	Yes	No

[a]GI: gastrointestinal absorption
[b]BBB: Blood–brain barrier
[c]P-gp: P-glycoprotein

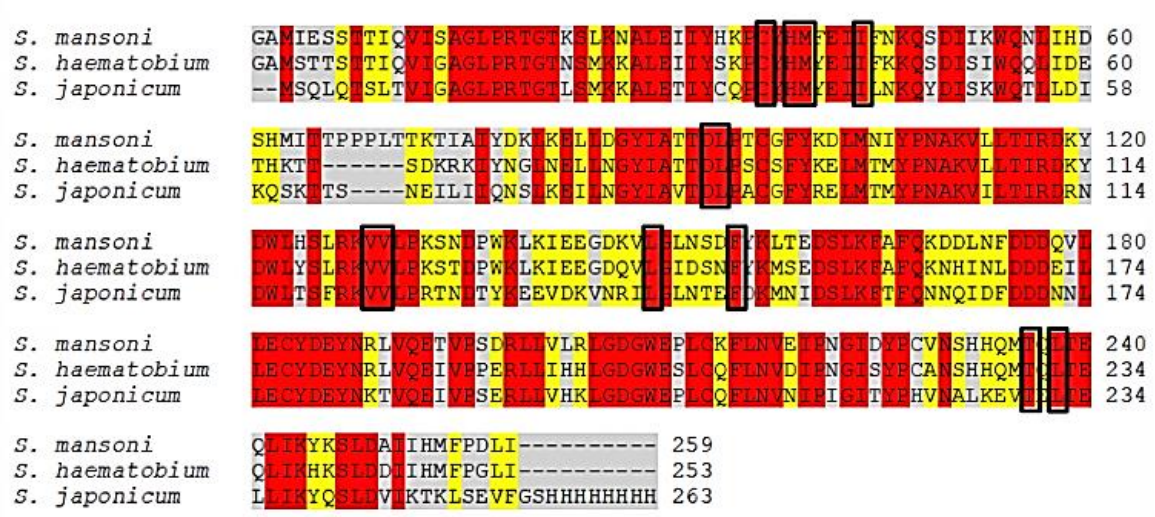
3.2 Alignment of protein sequences

We investigated the amino acids shared between the SULT sequences of the species *S. mansoni*, *S. haematobium*, and *S. japonicum* and observed that most residues of binding to the inhibitor OXA in *S. mansoni* are conserved in the other two species. Among the shared amino acids are: C35, H37, M38, I42, D91, L92 V127, V128, L147, F153, T237, and L240 (Figure 01). Other eight amino acids responsible for interaction with OXA in *S. mansoni* differ in *S. haematobium* and *S. japonicum* (F39, I140, G143, D144, L149, T157, M233, and

L256). Although the results show high similarity and identity between sequences of all the three species, it has been experimentally proven that OXA is sensitive only to *S. mansoni* (Valentim et al., 2014). Differences in catalytic efficiency due to *Sm*SULT, *Sh*SULT, and *Sj*SULT sulfation kinetics contribute to the difference in drug inhibitory activity (Taylor et al., 2017).

The human SULT (*h*SULT) is homodimeric, unlike species of the *Schistosome* genus that have only one protein chain. In addition, the strands are structurally different from *Schistosome* spp. and share only one residue from the active site. Thus, it is possible to develop selective drugs against Schistosomiasis without prejudice to the host.

Figure 1. Alignment of SULT protein sequences in *Schistosoma* spp. Gray regions correspond to non-similar and non-identical amino acids; red regions correspond to identical amino acids; yellow regions correspond to similar amino acids. Black boxes are regions of the active site and inhibitor binding conserved in *S. mansoni*, *S. haematobium*, and *S. japonicum*.



3.3 Consensus Docking

Selected 28 molecules were docked using five different scoring functions against the *S. mansoni* SULT target. The results showed that 17 molecules (8, 20, 31, 32, 57, 58, 59, 60, 83, 84, 86, 87, 88, 90, 91, 121, and 131) obtained satisfactory energy values in at least three docking programs (Table 2).

Interaction of compounds with the minimal energy of binding (57, 60, and 121) are demonstrated on Figure 2. Hydrophobic interactions were observed with the amino acids

Arg17, Asp91, Val127, Lys139, Ile140, Asp144, Thr157, Asn228, Met 233, and Thr237. Van der Waals interactions were also formed, mainly with molecule 121, which presented 16 interactions of this type. All amino acids that have interacted with lignans correspond to the region of the active site and most of them are also responsible for maintaining OXA binding in that region.

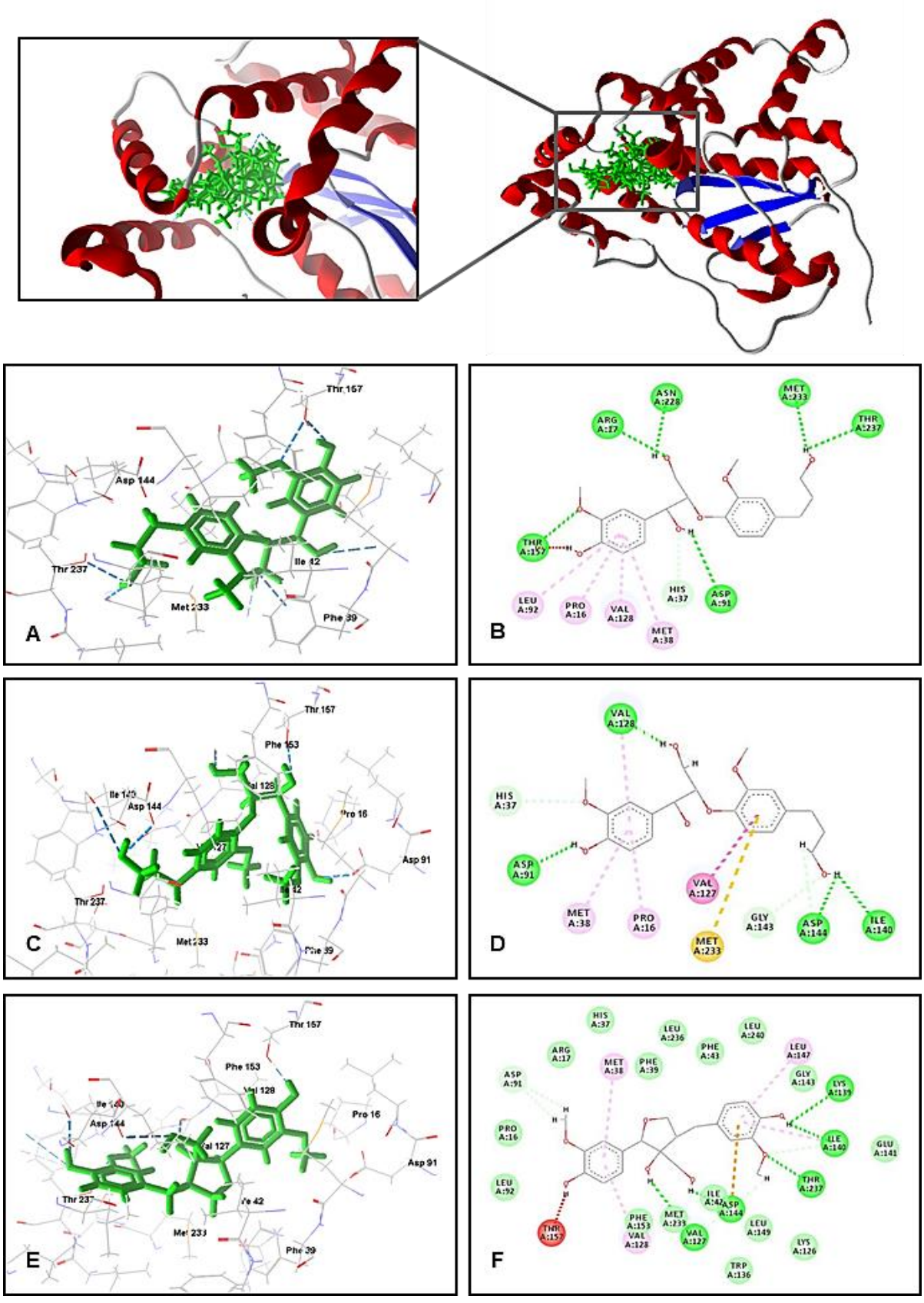
Table 2. Compounds with the highest scores in the consensus docking.

ID	MVD	GOLD	VINA	AD4	Plants
8	-81.73	41.21	-8.1	-8.28	45.83
20	-79.93	43.98	-9	-7.95	55.00
31	-87.27	48.75	-8.1	-6.95	61.01
32	-73.69	50.90	-8	-7.27	60.13
57	-90.45	54.05	-8.1	-5.79	69.66
58	-21.51	57.14	-7.8	-6.64	57.96
59	-79.30	50.06	-7.5	-6.02	56.59
60	-72.90	50.73	-8	-6.07	68.25
83	-3.04	48.19	-8.4	-8.04	52.58
84	-34.73	46.54	-8	-8.55	58.09
86	-83.70	42.29	-8.7	-7.49	54.58
87	-92.69	45.30	-8.1	-7.45	56.92
88	-82.30	42.80	-7.8	-6.42	45.91
90	-75.98	45.85	-6.5	-7.86	53.13
91	-77.07	0.00	-8	-6.88	57.17
121	-87.94	49.27	-8.1	-7.2	53.48
131	-60.89	33.76	-8.6	-7.31	53.21
OXA	-74.77	31.03	-7.5	-9.08	51.23

3.4 Prediction of biological activity

We have used two different techniques for QSAR modeling and virtual screening. The first model was generated using Random Forest (RF) and 128 Volsurf + 3D descriptors. Major statistical characteristics obtained for 5-fold external cross-validation were as high as follows: CCR = 0.87; Specificity = 0.91; Sensitivity = 0.82; Positive Predictive Value (PPV) = 0.86; Negative Predictive Value (NPV) = 0.87; ROC = 0.92; and Matthews coefficient (MCC) = 0.77.

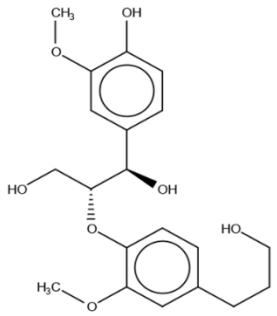
Figure 2. 2D and 3D interactions of the lignans (59 – A and B; 60 – C and D; 121 – E and F) with the SULT target. Van der Waals interactions are shown in light green, H-bonds – dark green, hydrophobic interactions - yellow, pink and red.



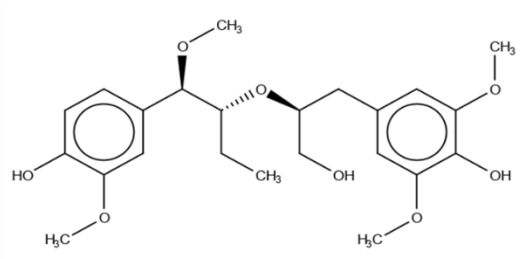
All 28 lignans selected before were within the AD. Only 3 compounds (44, 124, and 139) were predicted as actives, with their predictions varying between 53 and 57% probability. MuDRA was employed as a second approach. MuDRA has similar statistical characteristics to aforementioned RF model: CCR = 0.91; Specificity = 0.90; Sensitivity = 0.93; PPV = 0.88; NPV = 0.94; ROC = 0.95; MCC = 0.85. Seven lignans (44, 57-60, 124, and 132) were predicted as active against *Schistosoma mansoni* by MuDRA model. The MuDRA model selected 7 lignans with a probability of being active between 50 and 54% against *Schistosoma mansoni*. These lignans have chemical similarity varying between 0.6 and 0.7 with chemical structures that have schistosomicidal activity. MuDRA analysis for these seven compounds are depicted on Figure 3.

Through the consensus analysis of the RF and MuDRA model, it was possible to select seven lignans most likely to be active in the two virtual screening models generated from the ligand-based technique. Thus, compounds 57, 58, 59 and 60 were highlighted (Figure 4).

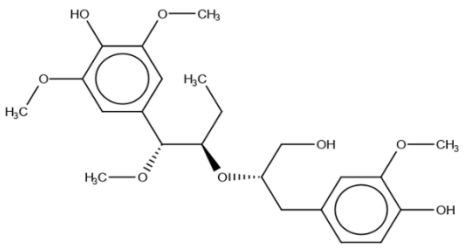
Figure 4. Molecules selected from a consensus model obtained by virtual screening models based on RF and MudRA ligands. The probability of biological activity based on the consensus model is indicated by Pc.



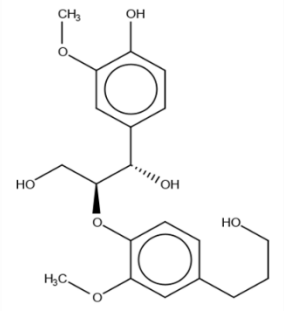
57
Pc = 0.65



58
Pc = 0.62

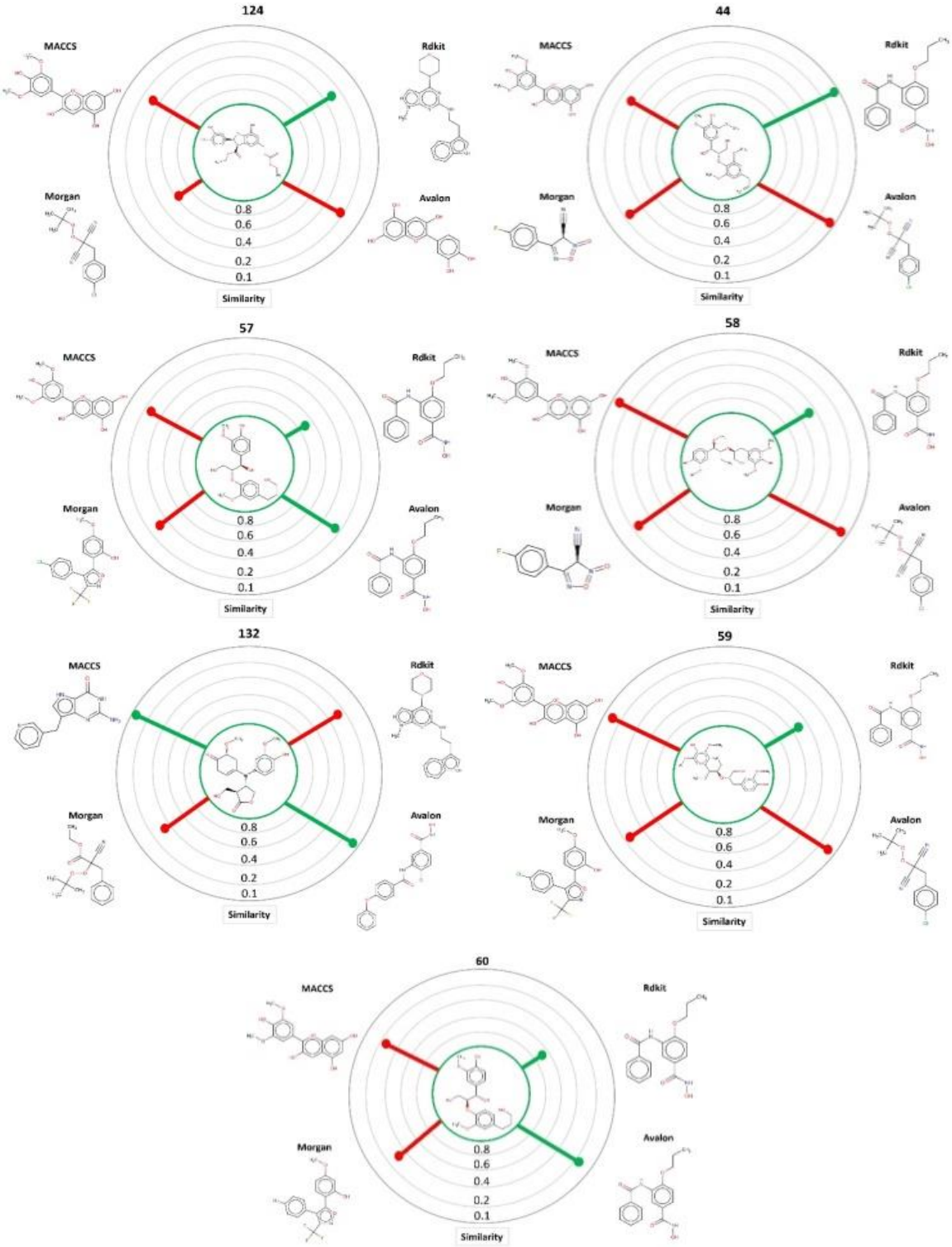


59
Pc = 0.59



60
Pc = 0.59

Figure 03. MuDRA plots for seven selected compounds. Green line – nearest neighbor is active against *S. mansoni*; Red - inactive. Green circle means that the compound is predicted to be active.



3.5 Molecular Dynamics Simulations

In order to characterize the interactions between SULT and the best binders, molecular dynamics (DM) simulations were performed. To study the stability and conformational changes of the complexes during the DM simulation, the mean square deviation (RMSD) was calculated for the protein atoms and the structures of each binder separately (Figure 04). The RMSD analysis of SULT showed that the protein reached conformations ranging from 0.2 to 0.35nm in size in 10ns, achieving stability from 2ns. The results show that the protein is stable, its conformation does not alter drastically. As for the binders, lignans 57, 58, and 60 are more stable, but compounds 58 and 60 have stability only for 4ns. We also observed that the OXA remained stable throughout the DM simulation.

To understand the flexibility of the residues and the amino acids that contribute to the conformational change of SULT, root mean square fluctuations (RMSF) of each amino acid of the protein were calculated. Residues with high RMSF values suggest more flexibility and low RMSF values reflect restrictions on the circulation of residues and thus have less flexibility (Zarezade et al., 2018). Considering that amino acids with fluctuations above 0.3nm contribute to the flexibility of the protein structure, we found that residues 1-7, 62, 63, 68-70, 222, 245-247, 249, and 254 are responsible for the change of the secondary structure (Figure 05). We also verified that none of the residues corresponds to the amino acids of the active site, which makes the region stable even after the interactions with the lignans and the OXA inhibitor.

The MM/PBSA method was used to further explore the arrangement of the interactions and to estimate the free energy after the DM simulation. As observed in Table 03, the OXA binder has binding-free energy higher than the lignans that obtained the best results in the docking and prediction of biological activity. Although lignans 57-60 showed higher affinity values than OXA, only MM/PBSA calculations could verify that in the presence of solvent, OXA performs better than the lignans. When analyzing the energy contributions, we noticed that the value of electrostatic energy favors the increase of free energy in OXA. High contribution of electrostatic energy in the binding of OXA to the active site of SULT could be explained by important role of Cys35 and that mutations in this residue may contribute to the resistance of the parasite to the drug (Valentim et al., 2014; Eng et al., 2017).

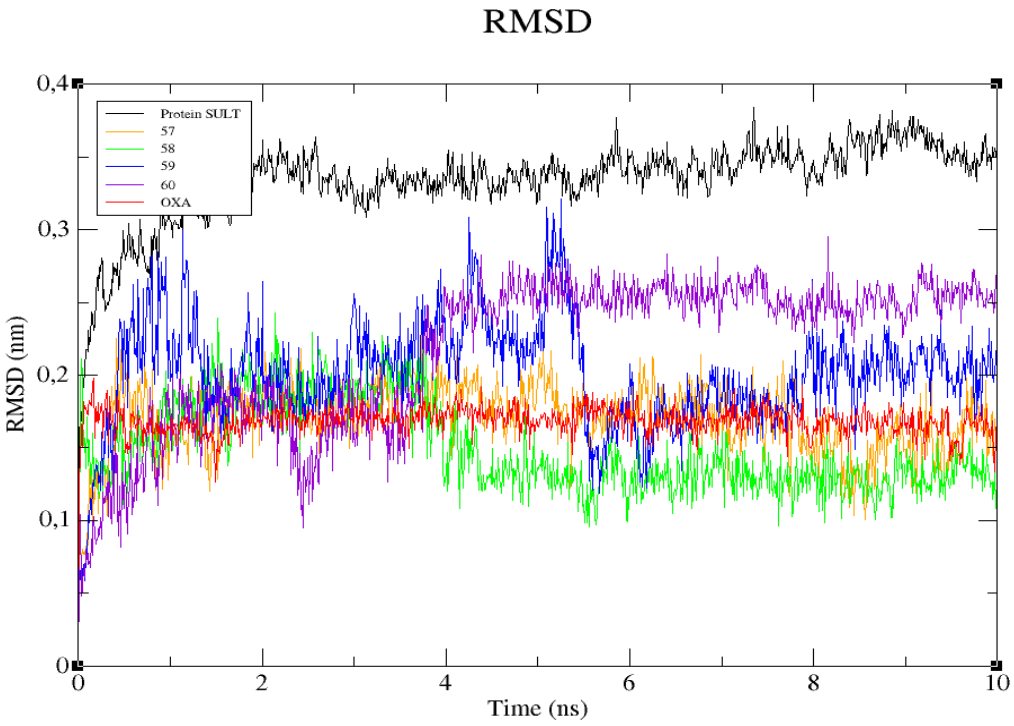


Figure 5. RMSD of the Cα atoms and ligands (lignans 57, 58, 59, 60 and PDB OXA).

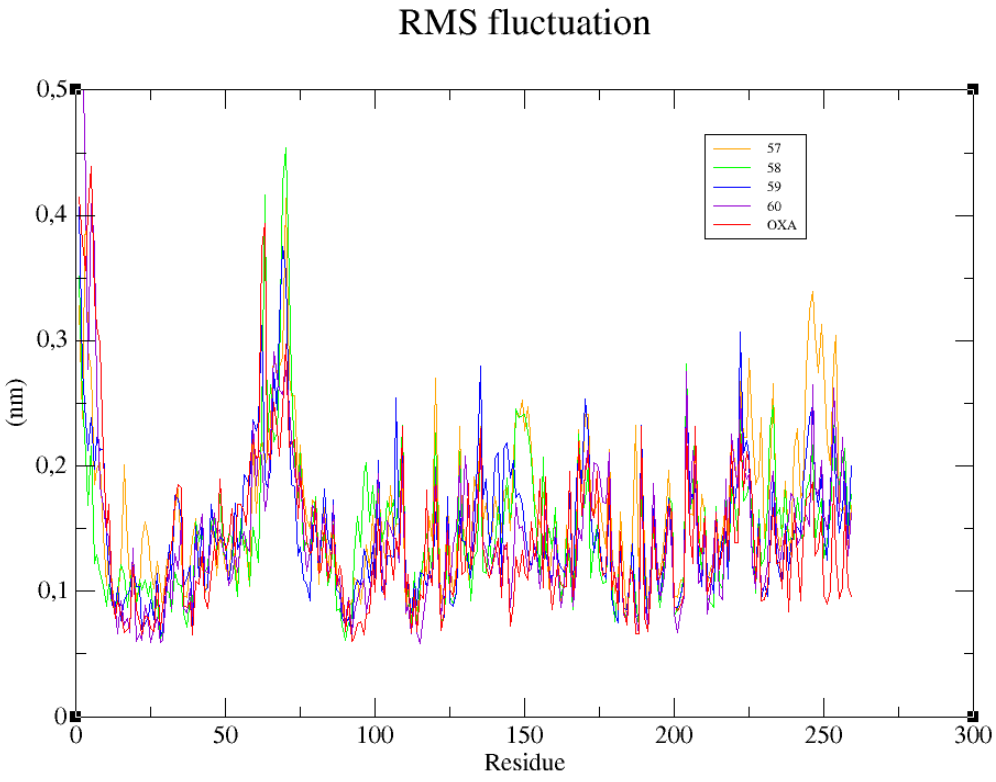


Figure 6. RMSF of the Cα atoms and ligands (Lignans 57, 58, 59, 60 and PDB OXA).

Table 03. Results of free energy calculations using the MMPBSA method.

ID	ΔE_{wdw}	$\Delta E_{Electrostatic}$	ΔG_{PB}	ΔG_{SA}	ΔG_{MMPBSA}
57	-231.06 ± 0.40	0.24 ± 0.19	47.40 ± 0.48	-21.08 ± 0.03	-204.49 ± 0.54
58	-251.26 ± 0.61	-14.87 ± 0.13	85.01 ± 0.57	-22.93 ± 0.04	-204.04 ± 0.59
59	-276.58 ± 0.59	-4.53 ± 0.12	49.69 ± 0.33	-23.19 ± 0.03	-254.59 ± 0.64
60	-246.24 ± 0.49	-7.29 ± 0.14	52.71 ± 0.29	-21.88 ± 0.03	-222.73 ± 0.57
OXA	-182.53 ± 0.53	-504.05 ± 1.15	408.63 ± 1.88	-16.73 ± 0.03	-294.71 ± 0.83

4. CONCLUSIONS

Schistosomiasis is one of the most widespread parasitic diseases in the world. Although two current medications on a market are effective, cases of reinfection and resistance have led researchers to discover new drugs against schistosomiasis. In order to obtain treatment alternatives from natural products, several in silico methodologies were used to screen lignans with a good ADMET profile and probability of potential therapeutic effect.

The enzyme Sulfotransferase is essential for the survival of the parasite and therefore we used it as a potential target. In this study, we were able to screen 28 lignans with high absorption, good solubility, good bioavailability, substrate for few CYP enzymes and low toxicity. Then, we obtained 17 molecules from consensus docking. Random Forest and MuDRA models were used to predict anti-parasitic activity and six molecules were considered satisfactory. When comparing the results of docking and QSAR, we observed that lignans 57-60 were scored high by both approaches. The results of the molecular dynamics simulation trajectories indicated that the SULT is stable and its active site also remains stable even complexed with the ligands. In the presence of solvent, free energy calculations have helped to understand that OXA behaves better than lignans. Even so, we observed through the interactions between SULT and lignans, that the amino acid Cys35 is not responsible for the stability of these compounds in the active site. It is important to note that this amino acid is responsible several cases of resistance have been attributed to this residue. Thus, we suggest that in cases of resistance, lignans are a great alternative for the treatment of Schistosomiasis.

Conflict of interest

The authors declare no conflicts of interest, financial or otherwise.

Acknowledgements

The present work was carried out with the support of the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

REFERENCES

Alves, V. M., Golbraikh, A., Capuzzi, S. J., Liu, K., Lam, W. I., Korn, D. R., Pozefsky, D., Andrade, C. H., Muratov, E. N., & Tropsha, A. (2018). Multi-Descriptor Read Across (MuDRA): A Simple and Transparent Approach for Developing Accurate Quantitative Structure-Activity Relationship Models [Research-article]. *Journal of Chemical Information and Modeling*, *58*(6), 1214–1223. https://doi.org/10.1021/acs.jcim.8b00124

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), 36–42. https://doi.org/10.1093/nar/gks1195

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1–3), 43–56.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1978). The protein data bank: A computer-based archival file for macromolecular structures. *Archives of Biochemistry and Biophysics*, *185*(2), 584–591. https://doi.org/10.1016/0003-9861(78)90204-7

Bondi., A. (1964). Van der Waals Volumes and Radii. *J. Phys.*, *68*, 441–451.

Cruciani, G., Pastor, M., & Guba, W. (2000). VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences*, *11*(SUPPL. 2). https://doi.org/10.1016/S0928-0987(00)00162-7

da Silva, V. B. R., Campos, B. R. K. L., de Oliveira, J. F., Decout, J. L., & do Carmo Alves de Lima, M. (2017). Medicinal chemistry of antischistosomal drugs: Praziquantel and oxamniquine. *Bioorganic and Medicinal Chemistry*, *25*(13), 3259–3277. https://doi.org/10.1016/j.bmc.2017.04.031

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small

molecules. *Scientific Reports*, *7*(October 2016), 1–13. https://doi.org/10.1038/srep42717

Dos Santos Maia M; de Sousa NF; Rodrigues GCS; Monteiro AFM; Scotti MT; Scotti L. (2020). Lignans and Neolignans anti-tuberculosis identified by QSAR and Molecular Modeling. *Comb Chem High Throughput Screen.* https://doi.org/10.2174/1386207323666200226094940.

Eng, N., Rugel, A. R., Ramiro, R., Assis, D., Oliveira, G., Holloway, S. P., Cao, X., Hart, P. J., Loverde, P. T., Anderson, T. J. C., Antonio, S., Gerais, M., & Antonio, S. (2017). *HHS Public Access*. *46*(7), 417–424. https://doi.org/10.1016/j.ijpara.2016.03.006.Independent

Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, *50*(7), 1189–1204. https://doi.org/10.1021/ci100176x

Fourches, D., Muratov, E., & Tropsha, A. (2015). Curation of chemogenomics data. *Nature Chemical Biology*, *11*(8), 535. https://doi.org/10.1038/nchembio.1881

Fourches, D., Muratov, E., & Tropsha, A. (2016). Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *Journal of Chemical Information and Modeling*, *56*(7), 1243–1252. https://doi.org/10.1021/acs.jcim.6b00129

Kumari, R., Kumar, R., & Lynn, A. (2014). G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Leger, E., & Webster, J. P. (2017). Hybridizations within the Genus Schistosoma: Implications for evolution, epidemiology and control. *Parasitology*, *144*(1), 65–80. https://doi.org/10.1017/S0031182016001190

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *64*(SUPPL.), 4–17. https://doi.org/10.1016/j.addr.2012.09.019

Lorenzo, V. P., Filho, J. M. B., Scotti, L., & Scotti, M. T. (2015). Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B. *Brazilian Journal of Pharmacognosy*, *25*(6), 690–697. https://doi.org/10.1016/j.bjp.2015.08.005

Mäder, P., Rennar, G. A., Ventura, A. M. P., Grevelding, C. G., & Schlitzer, M. (2018).

Chemotherapy for Fighting Schistosomiasis: Past, Present and Future. *ChemMedChem*, *13*(22), 2374–2389. https://doi.org/10.1002/cmdc.201800572

Mandal, S., Moudgil, M., & Mandal, S. K. (2009). Rational drug design. *European Journal of Pharmacology*, *625*(1–3), 90–100. https://doi.org/10.1016/j.ejphar.2009.06.065

McManus, D. P., Dunne, D. W., Sacko, M., Utzinger, J., Vennervald, B. J., & Zhou, X. N. (2018). Schistosomiasis. *Nature Reviews Disease Primers*, *4*(1), 1–19. https://doi.org/10.1038/s41572-018-0013-8

Melo-Filho, C. C., Dantas, R. F., Braga, R. C., Neves, B. J., Senger, M. R., Valente, W. C. G., Rezende-Neto, J. M., Chaves, W. T., Muratov, E. N., Paveley, R. A., Furnham, N., Kamentsky, L., Carpenter, A. E., Silva-Junior, F. P., & Andrade, C. H. (2016). QSAR-Driven Discovery of Novel Chemical Scaffolds Active against Schistosoma mansoni. *Journal of Chemical Information and Modeling*, *56*(7), 1357–1372. https://doi.org/10.1021/acs.jcim.6b00055

Merrifield, M., Hotez, P. J., Beaumier, C. M., Gillespie, P., Strych, U., Hayward, T., & Bottazzi, M. E. (2016). Advancing a vaccine to prevent human schistosomiasis. *Vaccine*, *34*(26), 2988–2991. https://doi.org/10.1016/j.vaccine.2016.03.079

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2010). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256.AutoDock4

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Porcino, G. N., Antinarelli, L. M. R., Maia, A. C. R. G., Faria-Pinto, P., Taunay-Rodrigues, A., Zech Coelho, P. M., Nelson, D. L., Penido, M. L. O., Coimbra, E. S., & Vasconcelos, E. G. (2019). The alkylaminoalkanethiosulfuric acids exhibit in-vitro antileishmanial activity against Leishmania (Viannia) braziliensis: a new perspective for use of these schistosomicidal agents. *Journal of Pharmacy and Pharmacology*, *71*(12), 1784–1791. https://doi.org/10.1111/jphp.13163

Sargis Dallakyan; Arthur J. Olson. (2014). Small-Molecule Library Screening by Docking with PyRx. *Chemical Biology*, 243–250.

Schüttelkopf, A. W., & Van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, *60*(8), 1355–1363. https://doi.org/10.1107/S0907444904011679

Taylor, A. B., Roberts, K. M., Cao, X., Clark, N. E., Holloway, S. P., Donati, E., Polcaro, C. M., Pica-Mattoccia, L., Tarpley, R. S., McHardy, S. F., Cioli, D., LoVerde, P. T., Fitzpatrick, P. F., & Hart, P. J. (2017). Structural and enzymatic insights into species-specific resistance to schistosome parasite drug therapy. *Journal of Biological Chemistry*, *292*(27), 11154–11164. https://doi.org/10.1074/jbc.M116.766527

Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321. https://doi.org/10.1021/jm051197e

Tomasz Puzyn , Jerzy Leszczynski, M. T. D. C. (2009). *An interdisciplinary overview on recent advances in Quantitative Structure – Activity Relationships (QSAR) studies ▶ Presents both a review of new ideas about methodology and new areas of QSAR applications ▶ A simply wr*. *8*(0), 6221. https://www.amazon.co.uk/Recent-Advances-QSAR-Studies-Computational/dp/1402097824

Valentim, C. L. L., Cioli, D., Chevalier, F. D., Cao, X., Taylor, B., Holloway, S. P., Pica-mattoccia, L., Guidi, A., Tsai, I. J., Berriman, M., Carvalho-queiroz, C., & Almeida, M. (2014). Specific drug action in Schistosome parasites. *Science (New York, N.Y.)*, *342*(6164), 1385–1389. https://doi.org/10.1126/science.1243106.Genetic

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics*, *52*(4), 609–623. https://doi.org/10.1002/prot.10465

Zarezade, V., Abolghasemi, M., Rahim, F., Veisi, A., & Behbahani, M. (2018). In silico assessment of new progesterone receptor inhibitors using molecular dynamics: a new insight into breast cancer treatment. *Journal of Molecular Modeling*, *24*(12). https://doi.org/10.1007/s00894-018-3858-6

Zoni, A. C., Catalá, L., & Ault, S. K. (2016). Schistosomiasis Prevalence and Intensity of Infection in Latin America and the Caribbean Countries, 1942-2014: A Systematic Review in the Context of a Regional Elimination Goal. *PLoS Neglected Tropical Diseases*, *10*(3), 1–22. https://doi.org/10.1371/journal.pntd.0004493

Capítulo 5

Embora a doença de Alzheimer (DA) seja caracterizada pela geração e/ou acúmulo de peptídeos amiloidogênicos, a patogênese da DA está associada principalmente ao estresse oxidativo. Por isso, neste capítulo, foram aplicadas abordagens computacionais sob diversos alvos que estão direta ou indiretamente envolvidas na via oxidativa. As lignanas também foram utilizadas neste estudo porque são conhecidas como produtos naturais com potencial atividade antioxidante e antiinflamatório. Por isso, as lignanas foram escolhidas e submetidas a diversas ferramentas computacionais para investigação do seu potencial para o tratamento da DA.

Inicialmente, foram construídos sete modelos de enzimas que participam da via oxidativa e inflamatória em humanos, onde sua ativação ou inativação podem contribuir para o desenvolvimento da DA, segundo estudos realizados com pacientes acometidos pela doença.

O docking molecular também foi realizado em várias enzimas participantes da via de interesse nesse estudo e uma análise consenso dos resultados do QSAR e do acoplamento molecular foram realizados para obter lignanas potenciais e multitarget para vários alvos. Em seguida, uma avaliação das propriedades ADMET dos compostos potenciais foi utilizada para determinar as melhores estruturas.

Identification of new targets and the virtual screening of lignans against Alzheimer's disease

Mayara dos Santos Maia, [1] Gabriela Cristina Soares Rodrigues, [1] Natália Ferreira de Sousa, [1] Marcus Tullius Scotti, [1] Luciana Scotti, [1]* and Francisco Jaime B. Mendonça-Junior [2]

[1]Laboratory of Cheminformatics, Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, João Pessoa-PB, Brazil.
[2]Laboratory of Synthesis and Drug Delivery, State University of Paraíba, João Pessoa-PB, Brazil .

*Correspondence should be addressed to Luciana Scotti; luciana.scotti@gmail.com

Received 22 April 2020; Revised 22 June 2020; Accepted 17 July 2020; Published 17 August 2020.

Artigo publicado na revista: Oxidative Medicine and Cellular Longevity

DOI: doi.org/10.1155/2020/3098673

Fator de impacto: 6.543

ABSTRACT: Alzheimer's disease (AD) is characterized by the progressive disturbance in cognition and affects approximately 36 million people, worldwide. However, the drugs used to treat this disease are only moderately effective and o not alter the course of the neurodegenerative process. This is because the pathogenesis of AD is mainly associated with oxidative stress, and current drugs only target two enzymes involved in neurotransmission. Therefore, the present study sought to identify potential multitarget compounds for enzymes that are directly or indirectly involved in the oxidative pathway, with minimal side effects, for AD treatment. A set of 159 lignans were submitted to studies of QSAR and molecular docking. A combined analysis was performed, based on ligand and structure, followed by the prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. The results showed that the combined analysis was able to select 139 potentially active and multitarget lignans targeting two or more enzymes, among them c-Jun N-terminal kinase 3 (JNK-3), Protein Tyrosine Phosphatase 1B (PTP1B), Nicotinamide Adenine Dinucleotide Phosphate Oxidase 1 (NOX1), NADPH Quinone Oxidoreductase 1 (NQO1), Phosphodiesterase 5 (PDE5), Nuclear factor erythroid 2-related factor 2 (Nrf2), Cycloxygenase 2 (COX-2) and inducible Nitric Oxide Synthase (*i*NOS). The authors conclude that compounds (06) austrobailignan 6, (11) anolignan c, (19) 7-epi-virolin, (64) 6-[(2*R*,3*R*,4*R*,5*R*)-3,4-dimethyl-5-(3,4,5-trimethoxyphenyl)oxolan-2-yl]-4-methoxy-1,3 benzodioxole, (116) ococymosin and (135) mappiodoinin b have probability

confer neuroprotection and antioxidant activity and represent potential alternative AD treatment drugs or prototypes for the development of new drugs with anti-AD properties.

1. INTRODUCTION

Although Alzheimer's disease is a multifactorial disease [1, 2], it is characterized by the increased generation and/or accumulation of amyloidogenic peptides (particularly Aβ), which are derived from the proteolysis of APP [3]. The presence of senile plaques in the cerebral cortex is thought to result in the activation of inflammatory and neurotoxic processes, culminating in the production of NO, cytokines, and ROS [3-9]. This process contributes to neurodegeneration and the loss of neuronal cells in AD [10, 11].

ROS can have beneficial and negative effects on cellular functions, depending on their concentrations. Low concentrations of ROS can regulate cellular functions, through redox-dependent signaling and redox-dependent transcription factors [8, 9]. However, high concentrations of ROS can impair vital cell processes, causing damage to proteins, lipids, and DNA [10]. Therefore, a balance between the production and removal of ROS is essential for normal cellular functions. Homeostasis imbalances can result in oxidative stress and the subsequent development of pathological conditions [11]. Stress precedes Aβ deposition, tau hyperphosphorylation, and impaired cognitive function. Endogenous antioxidant systems decrease with aging, favoring the appearance of AD. Therefore, oxidative stress is at the heart of AD pathogenesis [12, 13].

Currently, drugs for the treatment of AD include donepezil, galantamine and rivastigmine, which are inhibitors of the enzyme acetylcholinesterase, while memantine is a non-competitive inhibitor drug against N-methyl-D-aspartate (NMDA) [14-16]. These inhibitors act on cholinergic receptors and glutamate, respectively. This is because the oxidative glutamate toxicity [13] which is an excitatory neurotransmitter in the central nervous system (CNS), is associated with AD [16]. The excess of glutamate causes the suppression of cysteine uptake by the x_c^- system, which subsequently causes the inhibition of glutathione synthesis (GSH), triggering the accumulation of ROS [17, 18]. In addition to this mechanism, the neurochemical impairment of cholinergic neurons in the Central Nervous System (CNS) can contribute to the pathology of AD [19]. Although these drugs represent the best pharmacological treatments available at the time of AD, they have a

relatively small average overall effect and do not alter the course of the underlying neurodegenerative process [19]. Probably, because AD is multifactorial and is related to several deregulated mechanisms, due to the activation or inactivation of several enzymes important for homeostasis.

Knowing that oxidative stress is the center of the pathogenesis of AD, oxidative defense mechanisms appear to be important targets for the development of new and promising AD drugs. The Kelch-like ECH-associated protein 1 (Keap1)/Nrf2/ARE pathway is one of the most potent defensive systems against oxidative stress [20]. In addition, cyclooxygenase-2 (COX-2), inducible nitric oxide synthase (*i*NOS), NADPH oxidase (NOX), lipoxygenase (LOX), c-Jun N-terminal kinase 3 (JNK-3), protein tyrosine phosphatase 1B (PTP1B), phosphodiesterase type 5 (PDE5), NADPH oxidase, sodium-glucose cotransporter (SGLT)1, SGLT2, and DJ-1 have been associated with the expression of anti-inflammatory mediators, neuroprotection, and ROS regulation, and therefore represent promising AD targets [21-29].

Natural products are an important alternative for AD treatment because they contain widely known and reported classes of molecules associated with antioxidant activities, especially polyphenol compounds [23]. Lignans are a class of polyphenol compounds, which, according to Barbosa Filho in Simões (1999) [24], are chemically characterized as dimers formed by the oxidative homocoupling of cinnamic alcohols or the coupling with cinnamic acids.

Drug design is an important strategy in the field of medicinal chemistry, which increasingly requires the use of modern tools to ensure the increased practicality and speed of obtaining results. For example, we often utilize *in silico* studies that seek to understand the properties between a ligand and its respective receptor [25].

c-Jun N-terminal kinases (JNKs)

JNKs represent a family of serine-threonine protein kinases that are encoded by 3 genes (JNK1, JNK2, and JNK3) [26]. JNK1 and JNK2 are ubiquitously expressed, whereas JNK3 is primarily expressed in the brain. JNKs are activated by phosphorylation (pJNK), through the activation of mitogen-activated protein (MAP) mitogen-activated protein kinase (MAPK2), by extracellular stimuli, such as ultraviolet light, cytokines, and Aβ peptides [27]. In addition, studies have indicated that JNK can be activated by stress and triggered by

harmful external stimuli, via the kinase cascade and oxidative stress, in patients with AD [21]. JNKs are associated with several important functions in the cell, such as inflammation, the regulation of gene expression, cell proliferation, and apoptosis. JNK3 has been implicated in the pathogenesis of AD because JNK3 phosphorylates amyloid precursor protein (APP), which increases the production of Aβ [27]. Due to its fundamental role in neurodegeneration, JNK pathway signaling has been a target for the design of pharmacological and potential therapeutic agents [28].

The activation of the JNK pathway depends on the coordinated interaction among the scaffold proteins that belong to the JNK activation complex, which is capable of mediating signal amplification, ensuring substrate specificity, and coordinating a signaling cascade [29]. Different stimuli can trigger JNK activation, including JNK interaction protein 1a (JIP1a) and JIP1b (also called IB1), JIP2, and JIP3 (initially called JSAP1) JNK-associated leucine zipper protein (JLP), and various SRC-homology 3 domain (SH3)-containing proteins. Substrates are activated by JNK phosphorylation, mediated by c-Jun, which in turn interact with JunB, JunD, c-Fos, and activating transcription factor (ATF), which constitute the transcription factor activator protein 1 (AP-1), which regulates the maturation of the cellular response to stress and modulates the signals that ultimately lead to the activation of caspases and proteins associated with apoptosis [31, 33].

Studies have found elevated levels of JNK-3 in the brains of living patients with AD compared to levels in controls and that inhibitors kinases, including JNK-3, are able to reduce the effects of neuronal injury induced by Aβ [28, 32, 34, 35].

Phosphodiesterases (PDEs)

PDEs represent a group of enzymes, consisting of 11 subtypes (PDE1-PDE11), that control the *c*AMP and *c*GMP hydrolysis rates [33]. Variant PDEs play specific roles in different physiological characteristics and pathological processes. Although most PDE isoforms are expressed in the brain (PDE1, PDE2, PDE3, PDE4, PDE5A, PDE7A, PDE7B, PDE8B, PDE9A, PD10A, and PDE11A), their levels of expression vary among regions [34]. For example, PDE5 and PDE1 are located in the cerebellum, but only in Purkinje neurons; PDE1B is located in subsets of Purkinje cells; PDE6 is restricted to the retina and pineal gland; PDE3B is expressed in proopiomelanocortin and neuropeptide neurons; PDE1 exhibits distribution patterns in the hippocampus, cerebral cortex, thalamus, and striatum

[35]; PDE2A is widely expressed in the brain, with the strongest expression in the cortex, striatum, and hippocampus; and PDE4 is widely expressed in the CNS [22].

PDEs can affect neuronal cell survival, and when PDES malfunction, they can play roles in neurodegenerative diseases, such as AD [23]. PDE5 produces anti-inflammatory and neuroprotective effects, increasing NOS expression and *c*GMP accumulation and activating the protein kinase G (PKG) signaling pathway, which plays an important role in the development of several neurodegenerative diseases, including AD, Parkinson's disease (PD), and multiple sclerosis (MS) [24].

During AD pathogenesis, PDE5 hydrolyzes *c*GMP, an important intracellular messenger that activates PKG, triggering a wide range of intracellular signals [25]. The cyclic regulation of AMP/*c*GMP plays a determining role in several memory-related processes because these molecules are critical secondary messengers in the brain that are specifically associated with the memory recovery processes [35]. The levels of these messengers are maintained by the balance between production, catalysis, by adenylyl cyclase and guanylyl cyclase, and degradation, which is mediated by PDEs [36]. PDE5 specifically hydrolyzes *c*GMP [33]. Therefore, PDE5 inhibitors act to increase the levels of *c*GMP in neurons. Age-associated decreases in *c*GMP levels have been related to increased PDE5 expression and activity and the accumulation of Aβ peptide, which inhibits the activation of the NO/*c*GMP pathway [23]. Many studies have shown that PDE5 inhibitors exhibit therapeutic effects on AD by stimulating NO/*c*GMP signaling. PDE5 inhibitors can trigger vasodilation in the brain, resulting in the increased or sustained activation of signaling pathways that impact neuroprotective processes [39]. Therefore, elevating *c*GMP levels through PDE5 inhibition represents an alternative strategy for improving the learning and memory functions of AD patients.

Protein Tyrosine Phosphatase 1B (PTP1B)

PTP1B is a member of the non-transmembrane phosphotyrosine phosphatase family [37] and is a regulator of several processes in the CNS, many of which are therapeutically relevant to AD. Increased PTPB1 activity is associated with insulin deficiency and signaling pathways that are impaired in AD [38]. In addition, increased PTP1B activity can be activated with endoplasmic reticulum neuroinflammation and stress, which are both associated with amyloidosis [39]. The neuroinflammatory response includes the activation

of innate immune cells in the brain (microglia), the infiltration of macrophages, and the release of inflammatory mediators, such as NO, cytokines, and chemokines, which are associated with the progression of neurodegenerative diseases [37]. Inflammatory processes and amyloid aggregates have been implicated in neuronal loss and cognitive decline. When activated, PTP1B suppresses many signaling pathways that activate GSK3 and are involved in neurodegeneration.

Trodusquemina is a highly selective PTP1B inhibitor that has been used for the intervention of diabetes and obesity in clinical trials and has been investigated for the selective inhibition of PTP1B in neurons. The results showed that trodusquemina was sufficient to improve spatial learning and memory deficits in hAPP-J20 mice and to prevent the loss of neurons in the hippocampus [40]. In another study, PTP1B expression was found to be regulated by inflammatory stimuli, and PTP1B promotes microglial activation and functions as a critical positive regulator of neuroinflammation [37]. Thus, the inhibition of PTP1B provides a new therapeutic strategy for neuroinflammatory and neurodegenerative diseases.

Nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (NOX)

NOX is the most studied ROS-generating system [6]. NOX family members are transmembrane proteins that utilize electrons from cytosolic NADPH to reduce oxygen, generating a superoxide anion [16]. Seven known isoforms, NOX1, NOX2, NOX3, NOX4, NOX5, DUOX1, and DUOX2, combine with several subunits to form active enzyme complexes [41, 42]. The only known function of these membrane proteins is the catalysis superoxide anion formation from hydrogen peroxide. Hydrogen peroxide easily permeates cell membranes and can directly damage cells by oxidizing deoxyribonucleic acid (DNA), proteins, and lipids [42].

NOX primarily functions to generate free radicals, and some isoforms can be over-regulated by a variety of neurodegenerative factors [42]. Studies have suggested that the genetic and pharmacological inhibition of NOX enzymes may reduce harmful aspects associated with brain injuries and neurodegenerative disorders, resulting in a neuroprotective effect [42]. In particular, the observed lack of benefits associated with various antioxidant strategies may be due to the ineffectiveness of antioxidant molecules *in vivo* or the concomitant attenuation of oxidant regulatory roles [41]. Shimohama et al. [43]

reported the translocation of p47phox and p67phox, which strongly suggested that NOX is activated in the AD brain.

Studies with NOX inhibitors exert neuroprotective effects against AD, due to anti-inflammatory properties, through the oligomeric Aβ (oAβ)-induced microglial proliferation and the production of pro-inflammatory factors, including ROS, NO, tumor necrosis factor (TNF)-α, and interleukin (IL)-1β [43-46].

NADPH Quinone Oxidoreductase 1 (NQO1)

NADPH quinone oxidoreductase 1 (NQO1) is a flavin adenine dinucleotide (FAD)-dependent cytoplasmic flavoprotein that catalyzes the reduction of two electrons from quinones, quinonimines, nitroaromatic naphthoquinones, and substituted by glutathione, dichlorophenolindophenol (DCPIP) dyes, and an NADPH as an electron donor [12]. Therefore, NQO1, plays a central role in monitoring cellular redox status, protecting against oxidative stress induced by a variety of metabolic situations [45], including the metabolism of quinones and other xenobiotics, through the following mechanisms: (i) functioning as a two-electron donor, to provide a derivation that competes with the formation of ROS; (ii) maintaining reduced coenzyme Q; and (iii) regulating the stress-activated kinase pathway [46].

According to Chhetri et al. [12], the inactivation of the detoxifying enzyme NQO1 has been linked to the progression of AD. Factors that alter NQO1 activity can include genetic predispositions, such as the C690T NQO1 polymorphism, advanced age, cigarette smoking, and various medications [12]. The early expression of NQO1 in astrocytes may reflect a partially protective neuronal cell antioxidant protection system that activates at the beginning of the disease process, whereas the late expression of NQO1 may indicate the delayed activation of this system, as a final attempt to prevent neuronal cell death [47].

The antioxidant activity of NQO1 is essential; however, further studies are necessary to determine whether it should be targeted in the treatment of AD.

Nuclear factor erythroid 2-related factor 2 (Nrf2)

Nrf2 is a transcription factor that facilitates adaptation and survival under stress by regulating the gene expression of different networks of cytoprotective proteins, including anti-inflammatory and antioxidant proteins and proteins that repair or remove damaged

macromolecules [48]. Nrf2 plays a crucial role in maintaining cellular redox homeostasis and regulating the production of ROS by mitochondria. Nrf2 affects changes in the mitochondrial membrane potential ($\Delta\psi m$), ATP synthesis, and lipid peroxidation, and Nrf2 activation under stress conditions or by growth factors can neutralize increases in ROS production by the mitochondria, contributing to neuroprotection [49, 50].

Nrf2 is a key regulator of the body's antioxidant response and is responsible for inducing the expression of genes that encode antioxidant proteins and enzymes, in addition to metabolism detoxification phase II enzymes, which is a critical mechanism associated with cell protection and survival. Nrf2 targets include HO-1, superoxide dismutase (SOD), catalase (CAT), NADPH, NQO1, GSH S transferase (GST), GSH reductase (GR), GSH peroxidase (GPx), thioredoxin (Trx), and glutamate-cysteine ligase (GCL) [51, 52].

In addition to mediating antioxidant and detoxification mechanisms, Nrf2 is responsible for modulating the expression of 200 genes associated with other cellular processes, including the inflammatory response, metabolic regulation, cell proliferation, senescence, and mitochondrial function [53, 54].

Recent studies have investigated the participation of Nrf2, in the mechanisms of apoptosis and neuroprotection associated with Alzheimer's disease and traumatic brain injury, as well as the reduction of the expression of EROs [55].

Sodium-glucose transport protein (SGLT)

Glucose transporters can be divided into two primary families: facilitative glucose transporters (GLUTs) and sodium-dependent glucose cotransporters (SGLTs) [55]. Five primary SGLT isoforms have been identified, SGLT1, SGLT2, SGLT3, SGLT4, and SGLT5; however, SGLT1 and SGLT2, in particular, are associated with the pathways involved in the cellular mechanisms of AD [56].

The SGLT1 isoform is encoded by the SLC5A1 gene and performs glucose transport through a secondary active transport mechanism that uses the Na+ gradient established by the Na+/ K+ ATPase pump. This receptor is primarily expressed in the intestine, trachea, heart, testicles, prostate, brain, and kidneys. SGLT1 is characterized as a metabotropic receptor, coupled to transmembrane G proteins, with a secondary structure consisting of 664 amino acid residues, arranged in 14 transmembrane helices with both the NH2 and COOH terminals facing the extracellular side of the plasma membrane. The receptor contains only one N-glycosylation site, at Asn248 [57-59].

The SGLT2 isoform is encoded by the SLC5A2 gene and is found in the kidneys, brain, liver, thyroid, muscle, and heart. The SGLT2 structure is highly similar to that for the SGLT1 receptor and appears to be involved in diabetes and kidney disease mechanisms [55].

Studies have demonstrated the involvement of the factor SGLT1 in Alzheimer's disease, as it is related to cellular mediators of vascular injury [344]. Its activation is associated with a reduction in the levels of epidermal growth factor (EGFR), as well as its expression can be linked to food and control of insulin release by inhibiting the enzymes α-amylase and α-glucosidase [60-62].

Fator DJ-1

DJ-1 protein acts as an oxidative stress sensor and eliminates peroxide by self-oxidation [62]. This receptor is also related to cancer pathogenesis and may act as a potential tumor marker [63, 64]. DJ-1 participates in several signaling pathways, including mitochondrial quality control and the reaction to oxidative stress. Cells with high levels of DJ-1 have been shown to be resistant to oxidative stress and neurotoxins, such as 6-OHDA, whereas lower levels of DJ-1 make cells vulnerable to oxidative stress [65, 66].

The DJ-1 receptor was reported to have anti-Parkinson’s disease activity, by Dolgacheva and collaborators [67]. The mechanisms addressed included the protection of dopaminergic neurons against neurodegeneration in Parkinson's disease. The authors stated that the wild-type DJ-1 receptor can act as an oxidative stress sensor and as an antioxidant. DJ-1 regulates transcription and protects mitochondria from oxidative stress, in addition to increasing uncoupling protein (UCP)4 and UCP5 levels, which are responsible for mitochondrial decoupling and the consequent decrease in mitochondrial membrane potential. DJ-1 also suppresses the production of EROS and acts on redox factors, such as NF-κB, which acts on anti-inflammatory factors [68].

Cycloxygenase (COX)

Prostaglandins (PGs) are produced by prostaglandin-endoperoxide via synthase/cyclooxygenase (COX), which plays important roles in the etiology and inflammation of autoimmune diseases. COX has 2 isoforms: COX-1, which is permanently expressed in most tissues and organs, and COX-2, which is an inflammation-inducible

enzyme that is essential during the inflammation process and in autoimmune disease [69-73]. In addition, COX-2 plays a significant role in aging and skin cancer. PGE2 is a fundamental product of the COX synthesis pathway [71].

COX-2, also known as prostaglandin H synthase 2 (PGHS-2), catalyzes the conversion from arachidonic acid and O_2 to PGs, which are important lipid mediators involved in numerous physiological aspects and pathophysiological processes. Under normal physiological conditions, COX-2 most often has a low-level of expression, but this gene is highly induced in response to inflammation [72-74]. COX-1 is a constitutive enzyme, responsible for maintaining a basic level of PGs, to maintain physiological homeostasis, such as gastrointestinal integrity [74, 75]. COX-1 and COX-2 catalyze the biosynthesis of prostaglandins, prostacyclins, and thromboxanes [354]. COX-1 and COX-2 share a very high degree of sequence identity and very similar active site topologies [76].

Neurodegenerative diseases, such as AD, are sometimes treated with non-steroidal anti-inflammatory drugs (NSAIDs), which target COX-1 and COX-2 [77].

Nitric oxide synthase (NOS)

NOS is formed by a group of three enzymes (*e*NOS, *n*NOS, and *i*NOS), which are responsible for the generation of nitric oxide (NO) from the amino acid *L*-arginine [78],[79]. NO is a free radical gas and is associated with several biological functions, playing key roles in the regulation of blood flow, blood pressure, and oxygen delivery [80-82].

NOS includes endothelial NOS (*e*NOS or NOS1) [82, 83], inducible NOS (*i*NOS or NOS2) and neuronal NOS (*n*NOS or NOS3) [369]. *e*NOS and *n*NOS are characteristically expressed, whereas *i*NOS expression is induced exclusively by appropriate stimuli, such as cytokines, TNF-α, infections, chronic inflammation, tumors, interferon γ, or hypoxia [84]. During *i*NOS induction, the production of large amounts of NO occurs, in contrast with the other two isoforms [80, 85].

The generalized expression of *i*NOS in the CNS is pathological is often observed during neurological diseases, such as multiple sclerosis, stroke, and Parkinson's disease [86]. In patients with AD, studies have shown that the number of *i*NOS-positive neurons significantly increases in the brain and is associated with neuronal damage [87].

*e*NOS acts directly on the NO formation rate and acts as a limiting enzyme for this process, based on its expression levels and biological activity [79, 88]. eNOS activity also

influences the maintenance of vascular and endothelial homeostasis [89-91], in addition to the structure and function of the vascular endothelium [91].

*n*NOS produces NO in both the CNS and the peripheral nervous system, where it acts as a neurotransmitter [92, 93]. Although *n*NOS is the enzyme responsible for NO synthesis in neurons, not all neurons express *n*NOS [94]. However, the excessive activation of *n*NOS can result in neuronal death due, to the harmful production of NO [95].

Lipoxygenases (LOXs)

LOXs are a group of dioxygenase enzymes that contain iron and catalyze the stereoselective addition of oxygen to arachidonic acid (AA), docosahexaenoic acid (DHA), and other polyunsaturated fatty acids (PUFAs) [96]. The basic nomenclature of LOXs (except LOX-3) is based on the position of oxygen insertion in a substrate [96, 97]. Five types of LOXs have been identified in mammals, referred to as 5-, 8-, 12-, and 15-LOX and LOX-3 [98, 99].

Although 5-LOX is known primarily as a modulator of oxidation and inflammation [100], according to Chu et al. [101], this pathway can directly influence the pathogenesis of AD. The 5-LOX-γ-secretase pathway acts on the formation of Aβ peptides and other molecular diseases, including neuroinflammation, synaptic integrity, and cognitive function, which can contribute to new treatments for AD and associated neurodegenerative problems. High levels of 5-LOX in the nuclear envelope are associated with the release of leukotrienes to attract inflammatory cells [102].

5-LOX is widely distributed in the CNS and has been shown to be positively regulated in the *postmortem* brain of patients with AD, playing a functional role in the pathogenesis [103], as well as its activation influencing synapses and memory impairment [104]. According to Estrada Valencia et al. [105], 5-LOX is a key enzyme for AD because it is involved in inflammatory responses and is expressed at higher levels in the hippocampi of AD patients compared with healthy adults [106].

Observing that several enzymes are directly and indirectly involved through oxidative stress mechanisms, and that their activation and inactivation can contribute to neuroprotection or disease progression, the objective of the research was to explore new targets through virtual screening of lignans to identify molecules with potential anti-AD [107, 108].

2. MATERIALS AND METHODS

2.1 Data collection and curation

Several enzymes with available biological activity and 3D structure data were selected and investigated in this study. Chemical compounds were selected with known activity against the following enzymes: JNK-3 (CHEMBL2637), PTP1B (CHEMBL335), NFR2 (CHEMBL1075094), NOX1 (CHEMBL1287628), PDE5 (CHEMBL1827), COX-2 (CHEMBL230), and *i*NOS (CHEMBL4EM1). These compounds were used in the bank of images used to construct predictive models (https://www.ebi.ac.uk/chembl/) [58]. The details of the banks can be found in Table 1. The compounds were classified based on the pIC_{50} [-log IC_{50} (mol/l)]. The IC_{50} value represents the concentration required for 50% inhibition. However, for the enzyme Nrf2, activation data was used because the activation of this protein would obtain the desired effect. In addition, 159 CHEMBL lignans (Table S1) were assessed by virtual screening to identify molecules with potential activity against enzymes involved in AD progression, according to the workflows presented by Fourches et al. [81]. Three-dimensional structures were generated by ChemaxonStandardiser v.18.17.0, (www.chemaxon.org).

Table 1. Set of molecules from the ChEMBL databases for each enzyme selected in the study.

Database	Active molecules	Inactive molecules	Total
JNK-3	580 ($pIC_{50} \geq 6.0$)	642 ($pIC_{50} < 6.0$)	1.222
PTP1B	1.446 ($pIC_{50} \geq 5.0$)	1.354 ($pIC_{50} < 5.0$)	2.800
NFR2	163 (Activity)	85 (No activity)	248
NOX1	85 ($pIC_{50} \geq 4.75$)	60 ($pIC_{50} < 4.75$)	145
PDE5	873 ($pIC_{50} \geq 7.0$)	869 ($pIC_{50} < 7.0$)	1.742
COX2	2.018 ($pIC_{50} \geq 5.50$)	1.702 ($pIC_{50} < 5.50$)	3.720
***i*NOS**	396 ($pIC_{50} \geq 5.50$)	367 ($pIC_{50} < 5.50$)	763

2.2 Quantitative structure-activity relationship (QSAR) Modeling

Knime 3.5.3 software (KNIME 3.5.3, Konstanz Information Miner Copyright, 2018, www.knime.org) was used to perform the analyses and to generate the *in silico* models. Given the success of our previous studies [111, 112] we opted to perform a 3D QSAR analysis for each bank of enzymes. All studied compounds with a solved chemical structure

were saved in special data file (SDF) format and imported into Dragon 7.0 software [85], to generate descriptors.

The banks of molecules and their calculated descriptors were imported from Dragon software, and the data were divided into a “Partitioning” tool, using the “Stratified sample” option, which separated the data into Training and Testing sets, which represented 80% and 20% of all compounds, respectively. The sets were randomly selected, but the proportions of active and inactive substances were maintained in both databases.

The Random Forest (RF) algorithm, using WEKA nodes [114], was used to build predictive models. The parameters selected for RF for all models were as follows: the total number of forests was 250, and 1 seed was used for the generation of random numbers. Cross-validation was performed to estimate the predictive power of the developed models.

The external performances of the selected models were analyzed for sensitivity (true-positive rate, or active rate), specificity (true-negative rate, or inactive rate), and accuracy (general predictability). In addition, the sensitivity and specificity of the receiver operating characters (ROC) curve were used because this describes actual performance more clearly than accuracy.

The models were also analyzed using the Matthews correlation coefficient (MCC), which can evaluate the model globally, based on the results obtained in the confusion matrix. The MCC is a correlation coefficient between the observed and predictive binary classifications, resulting in values between -1 and +1, where a coefficient of +1 represents a perfect prediction, 0 represents a random prediction, and -1 indicates the total disagreement between the prediction and the observation [115].

MCC can be calculated using the following formula:

$$MCC = \frac{VP\ x\ VN - FP\ x\ FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

where VP represents true positives, VN represents true negatives, FP represents false positives, and FN represents false negatives.

The applicability domain (APD) was used to analyze the compounds in the test sets, to evaluate whether the predictions are reliable. The APD is a theoretical chemical space that encompasses the model's descriptors and the modeled response, allowing the estimation of uncertainty when predicting the activity of a compound in the training set used during the development of the model. This technique is important for verifying the reliability of QSAR

models by comparing predicted values with observed values [396]. APD is calculated using the following formula:

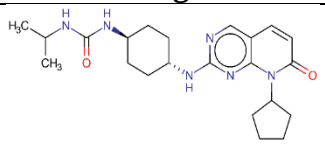
$$APD = d + Z\sigma$$

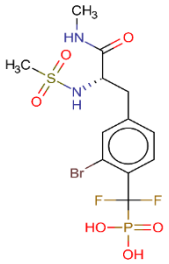
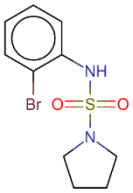
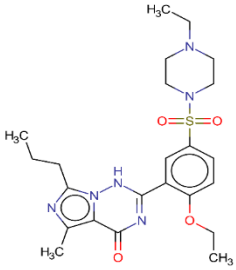
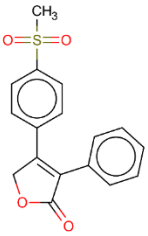
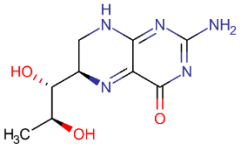
where d and σ are the Euclidean distances and the mean standard deviation, respectively, for the compounds in the training set. Z is an empirical cutoff value, which was set to 0.5 in this study [135].

2.3 Molecular docking

Molecular docking was performed using the Molegro Virtual Docker v6.0.1 (MVD) software [117], and six targets were selected for anchorage studies (Table 2). The 3D structures of the enzymes used in this study were obtained from Protein Data Bank (PDB) [118], using the following codes: PDB ID 4Y46 for JNK-3; PDB ID 4Y14 for PTPB1; PDB ID 6FY4 for NQO1; PDB ID 3B2R for PDE5; PDB ID 5KIR for COX-2; and PDB ID 4NOS for *i*NOS. We did not dock the enzymes Nrf2 and NOX1 because 3D structures were not available in PDB for the human species. Initially, all water molecules were removed from the crystalline structure, and the root-mean-square deviation (RMSD) was calculated from the poses, which indicates the degree of reliability for the fit. The RMSD provides for the connection mode close to the experimental structure and is considered successful if the value is below 2.0 Å. The MolDock score algorithm was used as a scoring function, to predict the best interactions between the ligand and the receptor. Then, the anchor assistant was generated, in which the enzyme and ligands were inserted to analyze the stability of the system based on the interactions identified with the active site of the enzyme.

Table 2. Information regarding the selected enzymes, obtained from the PDB database and used for docking.

PDB ID	Enzyme	Class	PDB ligand	Resolution
4Y46	c-Jun N-Terminal Kinase	Transferase		2.04 Å

4Y14	Tyrosine phosphatase 1B	Hydrolase		1.89 Å
6FY4	NAD(P)H:quinone oxidoreductase	Oxidoreductase		2.76 Å
3B2R	Phosphodiesterase-5	Hydrolase		2.07 Å
5KIR	Cyclooxygenase-2	Oxidoreductase		2.69 Å
4NOS	Inducible nitric oxide synthase	Oxidoreductase		2.25 Å

2.4 Prediction of ADMET properties

ADME parameters were calculated using the SwissADME open-access web tool (http://www.swissadme.ch) [119], which offers a set of rapid predictive models for the assessment of physicochemical, pharmacokinetic, and pharmacological properties. The toxicity prediction was performed in OSIRIS Property Explorer (https://www.organic-chemistry.org/prog/peo/) [120], based on the following parameters: mutagenicity, tumorigenicity, reproductive effects, and irritability. For absorption, factors included membrane permeability, intestinal absorption, and substrate or inhibitor of P glycoprotein. Thus, we investigated compounds that did not exceed more than two violations of Lipinski´s rule and for which the logP consensus was not greater than 4.15. In addition,

compounds were not substrates for the permeability glycoprotein enzyme (P-gp). The distribution was assessed by factors that include the blood-brain barrier (logBB) and the permeability of the CNS. Metabolism was predicted based on the CYP substrate or inhibition models (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4).

3. Results and Discussion

3.1 QSAR modelling

The metrics mentioned are the most commonly used metrics for chemoinformatics, although others that can be used to guarantee the high predictability of the model, such as ROC curves [121]. The results of the ROC curve and MCC analyses revealed excellent results. The models achieved ROC curves greater than 0.78 during cross-validation, and the MCC values were also greater than 0.52 during the cross-validation, revealing a model with excellent classification, performance, and robustness (Table 3, Figure S1). Only the model for the Nrf2 enzyme achieved an MCC below 0.5. Table 4 shows the ROC curve values for each protein.

Table 3. Performance summary corresponding with the results obtained for all Random Forest models.

Enzyme	Validation	Accuracy	Sensitivity	Specificity	PPV	NPV	MCC
JNK-3	Test	0.89	0.91	0.87	0.86	0.91	0.78
	Cross	0.83	0.85	0.82	0.81	0.85	0.67
PTP1B	Test	0.81	0.81	0.81	0.82	0.80	0.62
	Cross	0.82	0.82	0.82	0.83	0.81	0.64
NFR2	Test	0.76	0.75	0.76	0.86	0.61	0.50
	Cross	0.73	0.78	0.63	0.80	0.60	0.41
NOX1	Test	0.82	0.76	0.91	0.92	0.73	0.67
	Cross	0.80	0.89	0.66	0.92	0.73	0.58
PDE5	Test	0.87	0.9	0.84	0.85	0.9	0.75
	Cross	0.86	0.88	0.85	0.85	0.87	0.73
COX2	Test	0.78	0.83	0.71	0.77	0.78	0.55
	Cross	0.76	0.81	0.7	0.76	0.76	0.52
***i*NOS**	Test	0.81	0.87	0.74	0.78	0.84	0.62
	Cross	0.8	0.85	0.74	0.78	0.82	0.60

Table 4. Values for the ROC curves, during the test and cross-validation, for each RF model.

Enzyme	ROC curve	
	Test	Cross
JNK-3	0.96	0.91
PTP1B	0.87	0.89
NFR2	0.82	0.81
NOX1	0.90	0.78
PDE5	0.95	0.94
COX2	0.84	0.84
***i*NOS**	0.87	0.87

Using the models created, with excellent performance, the lignan set was screened to select compounds that are potentially active against the studied enzymes. Lignans with a probability of biological activity above 0.5 and that passed the applicability domain were considered active.

The results showed that no lignans were considered active for the JNK-3, PDE5 and COX-2 targets. However, 22 compounds were potentially active against the PTPB1 enzyme with a probability ranging from 50-74%, 111 compounds active against Nfr2 with a probability ranging from 50-64%, six compounds active against NOX1 with a probability ranging between 63-78% and 27 compounds active against *i*NOS with probability varying between 52-79%.

3.2 Docking molecular

The molecular docking study was performed for six enzymes that were targeted for the AD treatment. The lignan set was analyzed to select molecules with good probabilities for potential inactivation and activation activity against the enzymes targeted for AD treatment. Docking was not performed for Nrf2 and NOX1, due to the unavailability of human 3D protein structures.

In this study, the docking results were validated by the re-docking of the crystallographic ligand and by the RMSD of the poses. Re-docking consists of positioning and predicting the binding affinity of the crystallographic ligand in the region of the active site of the enzyme. The RMSD compares and calculates the mean deviation of the square root of the poses obtained by re-docking and the structure of the ligand obtained experimentally. For the fit to be reliable, the RMSD value must be 2.0 Å or less. The results

showed that the targets JNK-3, PTP1B, NQO1, PDE5, COX-2 and *i*NOS obtained RMSD values of 0.56, 0.25, 0.18, 0.47, 0.19 and 0.16 Å, respectively.

Molegro software is capable of generating interaction energies for lignans, by producing a MoldockScore for each studied protein. Then, calculations were performed to identify the lignans with the best active potential probabilities for each analyzed protein, using the following formula:

$$Prob = \frac{E_{Lig}}{E_{MLig}}, se\ E_{Lig} < E_{Inib}$$

where E_{Lig} is the energy of the analyzed lignan, E_{MLig} is the lowest energy obtained from the tested lignans, and E_{Inib} is the energy of the inhibitor ligand, obtained from the crystallography data for the tested protein. Only molecules that obtained binding energies below the binding energy for the crystallographic inhibitor ligand were considered to be potentially active.

Table 5 shows the interaction energies of the inhibitor ligand for each protein, and the top ten lignans with the best energy values for each protein.

Table 5. MoldockScore scores for the top ten lignans with the best energy values relative to the energy value of the crystallographic ligand for each protein.

ID	JNK-3	PTP1B	NQO1	PDE5	COX2	*i*NOS
1	-183	-177	-137	-204	-203	-178
2	-175	-156	-137	-192	-193	-153
3	-164	-154	-136	-182	-191	-147
4	-159	-153	-124	-169	-190	-144
5	-155	-153	-120	-167	-187	-143
6	-148	-152	-116	-166	-176	-143
7	-148	-152	-116	-164	-175	-143
8	-146	-151	-114	-164	-174	-141
9	-146	-151	-112	-164	-172	-139
10	-144	-150	-108	-162	-170	-139
Ligand PDB	-134	-156	-36	-139	-142	-59

Among the 159 lignans analyzed by molecular docking, 21 were found to be potentially active against JNK-3, 1 was identified for PTP1B, 157 were identified for NQO1, 34 were identified for PDE5, 53 were identified for COX-2, and 156 were identified for *i*NOS. These results indicated that lignans, in general, are more likely to activate the NQO1 and *i*NOS proteins, and are not selective for the PTP1B enzyme.

Among the 159 lignans analyzed by molecular docking, 21 were found to be potentially active against JNK-3, 1 was identified for PTP1B, 157 were identified for NQO1, 34 were identified for PDE5, 53 were identified for COX-2, and 156 were identified for *i*NOS. These results indicated that lignans, in general, are more likely to activate the NQO1 and *i*NOS proteins, and are not selective for the PTP1B enzyme.

3.3 Combined analysis based on ligand and structure

A second consensus analysis was performed to identify potential multitarget lignans, which demonstrate active potential probabilities for more than one protein, based on the RF model and docking. In this case, we use all the results of prediction of biological activity of the lignans and combine them with the results of docking. For this analysis, the following formula was used:

$$Prob_{Comb} = \frac{(ProbDc + (1 + ESP) \times P_{Activity}}{2 + ESP}, Se\ Prob_{Comb} > 0.5$$

where $Prob_{Dc}$ is the active potential probability from the molecular coupling analysis, *ESP* is the average specific value of the RF model, and $P_{Activity}$ is the active potential probability value of the RF model. This combined probability was conditioned, as only molecules with values greater than 0.5 were considered likely to be active. Combined probability values were calculated for the lignans identified for each target enzyme, and we analyzed which molecules were multitarget.

After performing the combined analysis, based on the ligand and structure, and using the formula to identify multitarget molecules, we identified 139 molecules that were potentially active for two or five target enzymes, out of the entire lignan set analyzed. For Nrf2 and NOX1, we only used the biological activity probability data, and for NQO1 we only used the docking data not enough data was available for these enzymes to construct the necessary models.

The combined probability ($Prob_{Comb}$), based on both ligand and structure, can increase the predictive power of the models and decrease the number of false positives. Combined probability analyses could be performed for five enzymes (JNK-3, PTP1B, PDE5, COX-2, and *i*NOS). For enzymes without sufficient data to build both models, only model was used. For molecules to be considered potentially active, the probability values

should be equal to or greater than 0.5. However, for $Prob_{Dc}$, the probability value should also be greater than that for the crystallographic ligand.

After the combined probability analysis, we selected the multitarget compounds that passed the applicability domain for all enzymes in this study. Using $Prob_{Comb}$, we were able to select three compounds with probabilities of activity ranging from 50%-61% for JNK-3, 43 compounds with a 52-72% probability for PTP1B, 57 compounds with 51%–72% probabilities for PDE5, 27 compounds with probabilities between 50%–61% for COX-2, and 27 compounds with probabilities between 50%–81% for *i*NOS (Table 6). The number of compounds with excellent combined probabilities was reduced when compared with the results of the docking probabilities; however, the combined probabilities increased the numbers of true positives.

Based on the biological activity probability data, 111 compounds, with probabilities ranging from 50%–64%, were identified for Nrf2, and nine compounds, with probabilities ranging from 51%–78%, were identified for NOX1. Based on the docking probability data, 156 compounds were selected, with probabilities ranging from 27%–100%, for NQO1. For this enzyme, compounds with probabilities above 0.27 were considered, as these were greater than the probability of the crystallographic ligand, which was 0.26.

We observed that although the results of QSAR do not indicate active compounds for JNK-3, PDE5 and COX-2, after the application of the formula that combines prediction values of biological activity and docking ($Prob_{Comb}$), we were able to identify active compounds for all targets of the study.

Table 6. Potentially active lignans, multitarget for four or more enzymes, based on the RF and docking model. In bold, are highlighted the active enzymes that walk in the applicability domain.

ID	$Prob_{Comb}$					$Prob_{Activity}$		$Prob_{Dc}$	Multitarget
	JNK-3	PTP1B	PDE5	COX-2	*i*NOS	NFR2	NOX1	NQO1	
05	0.39	**0.68**	**0.52**	0.41	**0.62**	**0.54**	0.17	**0.47**	5
06	0.38	**0.67**	0.49	0.43	0.59	**0.57**	**0.51**	**0.35**	4
07	0.45	**0.66**	**0.56**	0.53	0.70	**0.59**	0.25	**0.49**	4
11	0.35	**0.64**	0.48	0.46	0.53	**0.53**	**0.63**	**0.38**	4
12	0.37	**0.62**	**0.51**	0.48	0.60	**0.60**	0.25	**0.64**	4
13	0.32	**0.62**	**0.59**	0.46	0.57	**0.51**	0.45	**0.72**	4
14	0.51	**0.62**	**0.54**	0.45	**0.62**	**0.60**	0.25	**0.45**	5
19	0.31	**0.59**	0.48	0.49	0.56	**0.56**	**0.51**	**0.35**	4
33	0.41	**0.53**	**0.51**	**0.50**	0.51	**0.58**	0.41	**0.39**	5

34	0.35	**0.53**	**0.50**	0.43	**0.54**	**0.56**	0.45	**0.40**	5
35	0.27	**0.52**	**0.59**	0.46	0.49	**0.61**	**0.70**	**0.56**	5
38	0.54	**0.52**	**0.69**	**0.51**	0.66	**0.52**	0.31	**1.00**	5
39	0.54	**0.52**	0.68	**0.61**	0.65	**0.61**	0.33	**0.81**	4
41	0.59	**0.52**	0.61	**0.60**	0.54	**0.58**	0.33	**0.67**	4
42	0.52	**0.52**	0.64	**0.59**	0.62	**0.57**	0.36	**0.78**	4
44	0.35	**0.51**	**0.52**	0.45	0.58	**0.54**	0.43	**0.50**	4
45	0.45	**0.51**	**0.52**	0.49	0.55	**0.54**	0.35	**0.47**	4
47	0.39	**0.50**	**0.56**	0.47	0.54	**0.56**	0.36	**0.56**	4
52	0.45	0.48	0.49	**0.50**	**0.69**	**0.54**	0.45	**0.38**	4
104	0.42	0.40	**0.55**	**0.51**	0.31	**0.59**	0.30	**0.50**	4
106	0.48	0.40	**0.58**	**0.57**	0.52	**0.56**	0.25	**0.63**	4
108	**0.53**	0.40	**0.56**	0.45	0.64	**0.50**	0.38	**0.66**	4
115	0.31	0.39	**0.52**	**0.51**	0.61	**0.52**	0.28	**0.52**	4
134	0.26	0.36	**0.50**	0.40	**0.51**	**0.56**	**0.71**	**0.38**	5
141	0.40	0.35	**0.54**	**0.50**	0.50	**0.62**	**0.52**	**0.40**	5
142	0.37	0.35	**0.50**	**0.51**	0.65	**0.55**	0.30	**0.40**	4
146	0.40	0.33	**0.57**	**0.56**	0.66	**0.55**	0.27	**0.60**	4
153	0.49	0.32	**0.60**	**0.50**	0.63	**0.52**	0.39	**0.74**	4

3.4 Prediction of ADMET properties

The set of 139 potentially active and multitarget lignans were submitted to several predictive parameters to identify the compounds with the best ADMET profiles. Using physical-chemical properties, we attempted to verify compounds with good absorption, considering the lipid rule as a parameter.

According to Lipinski and Weber [43, 44], molecules with molecular weights below 500 Da, calculated LogP (CLogP) values less than five, less than five hydrogen bond donors, no more than ten hydrogen bond acceptors, and ≤ 10 rotating bonds have excellent absorption and bioavailability. Molecules that violate two or more of these rules do not show good absorption. We observed that 66% (92) of our lignans set showed solubility values that varied between soluble and moderately soluble.

Factors such as lipophilicity and solubility contribute to drug distribution *in vivo,* which is a requirement for advancing to preclinical and clinical tests. The most common descriptor for lipophilicity is the partition coefficient between *n*-octanol:water (log P). Ideal logP values are below 5.0. The results showed that 87% (121) of our lignan compounds had ideal logP values.

Metabolism can affect drug activity by changing the half-life, promoting the generation of toxic metabolites, or disrupting therapeutic potential. Pharmacokinetics are

essential for understanding drug metabolism in the body. For a compound to display the desired effect during AD treatment, the drug must be able to cross the blood-brain barrier. Many compounds that have been developed fail at the preclinical and clinical testing stage due to metabolism effects and poor absorption in the brain. Currently, the prediction and selection of compounds that act on nervous system tissues can be performed through *in silico* tests. The results showed that among lignans that target three or more enzymes, nine lignans would likely cross the blood-brain barrier.

Toxicity was also evaluated, and among the compounds that appeared likely to cross the blood-brain barrier, compounds 6, 11, 19, 64, 116, and 135 had no predicted mutagenicity or tumorigenesis effects, or negative effects on the reproductive system and irritability. Therefore, these molecules were considered to have the best ADMET properties because they do not present any toxicity risks. Tables S2 and S3 show the ADMET profiles of compounds with potential activity and multitargeting effects against four or more enzymes. In addition, Table S4 and Figure 1 show the compounds that did not present toxicity for these evaluated parameters.

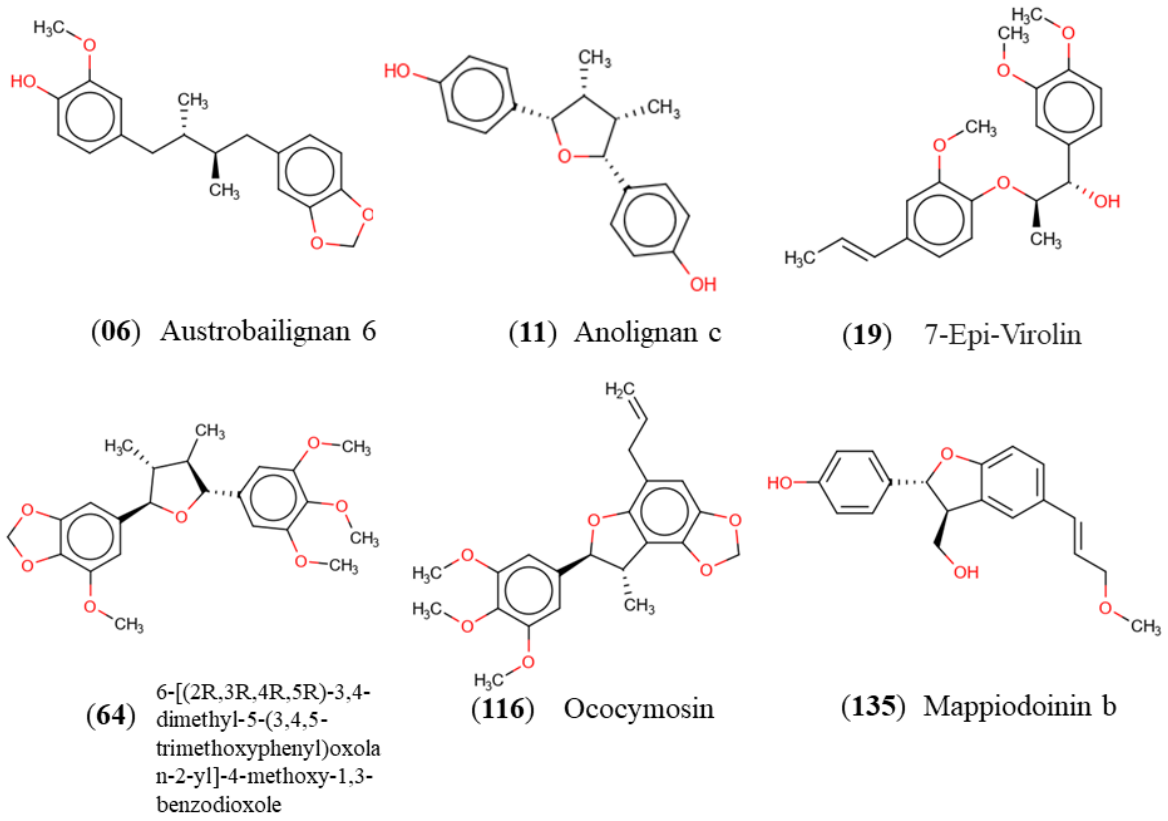


Figure 1. Lignans considered to be potentially active according to the Random Forest model, with multitarget effects and no predicted toxicity.

Due to the antioxidant properties of lignans, the present study sought to perform a virtual screening among diverse structural lignans to identify potential molecules for the treatment of AD. Lignans represent a huge class of pharmacologically active compounds that exhibit various functionalities, which are worth exploring by pharmaceutical industries [122].

According to a review by Zálešák et al. [123], several researchers have identified the antioxidant activity and neuroprotective properties of lignans. Lignans isolated from *Schisandra bicolor var*. were assayed for their neuroprotective effects against SH-SY5Y cell damage induced by Aβ25−35. Among the active compounds, both new lignans [esquibitubina B (L1-4), F (L1-7), H (L4-1) and I] and previously isolated lignans [galgravine, (-)-nectandrin A, (-)-futocadsurine A, (+)-9′-hydroxigalbelgin, austrobailignan-6, oleiferin-F, (+)-dihydro-guaiaretic acid, and (-)- isootobafenol] increased the cell viability in SH-SY5Y cells, following the induction of cellular injury by 3.25 nM Aβ25-35 compared with the negative control group. Furthermore, 25 μM dibenzocyclooctadiene lignans (L6-14 and NL5-10) from *Schisandra chinensis* exhibited protective activity against Aβ1-42 neurotoxicity induced in PC12 cells, increasing cell viability to 84.1% ± 5.4% and 82.1% ± 4.3%, respectively, compared with the control (52.0% ± 3.2%) [123].

Lignans are a large group of naturally occurring phenols widespread in the plant kingdom. In addition, notable advances have been made in the isolation and identification of lignans the last few years, which has already led to around 500 new congeners [122]. In addition, several studies have reported the synthesis of different lignans successfully and which have been tested for various biological activities [133-137].

3.5 Interaction analysis

We analyzed the interactions of six lignans through molecular docking that obtained the highest probability of activity, multitarget and with low toxicity. In addition, we consider analyzing the targets on which these compounds were most active.

The compounds austrobailignan 6 (06), anolignan c (11) and 7-Epi-virolin (19) formed several interactions with the PTP1B active site. Austrobailignan 6 formed hydrophobic interactions with residues Ile219 and Arg221, steric interactions with the amino acids Phe182, Cys215 and Ala217, an electrostatic bond with Arg181 and a hydrogen bond with Tyr46, stabilizing the bond. Analignan c formed four hydrophobic interactions with the amino acids Tyr26, Cys215, Ala217 and Arg221. In addition, it formed

an electrostatic and a steric interaction. 7-Epi-viroline formed several hydrophobic interactions with Tyr46, Phe182, Ala217 and Arg221. Three important hydrogen bonds were also observed with the residues Arg47, Arg45 and Glu262 (Figure 2).

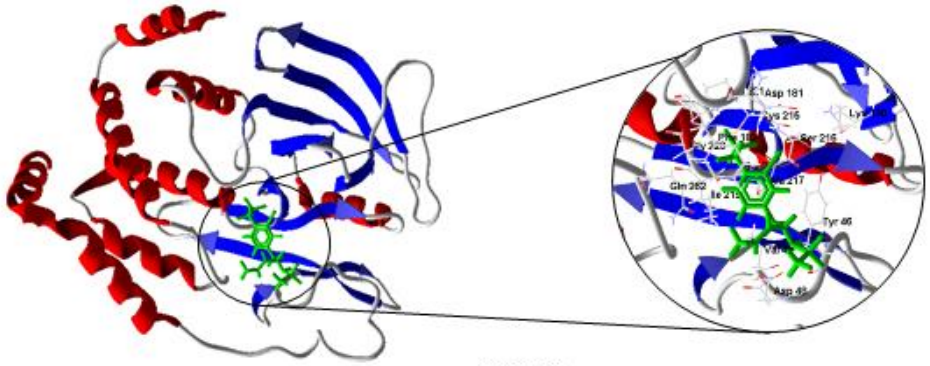
According to the study carried out by Krishnan et al.[126], the inhibitor CPT157633 managed to form electrostatic interactions with the PTP1B active site. In that study, interactions with the amino acids Cys215, Arg221 and Gln262 were reported. We observed that these amino acids are also interacting with lignans, forming more stable bonds.

These same lignans were also investigated for their interactions with the NQO1 target. We found that 6 - [(2R, 3R, 4R, 5R) -3,4-dimethyl-5- (3,4,5-trimethoxyphenyl) oxolan-2-yl] -4-methoxy-1,3-benzodioxole (64) formed hydrogen bonds with the amino acids Tyr129, Gly175 and Ile176, and a hydrophobic interaction with the amino acid Tyr127. Oocymosin (116) showed hydrophobic interactions with Tyr127 and Phe179. In addition, it formed a hydrogen bond with the Tyr129 residue. Mappiodionin b (135) formed hydrogen bonds with Gly175 and Ile176 and a hydrophobic interaction with Tyr127. All compounds formed interactions with the same amino acids (Figure 3).

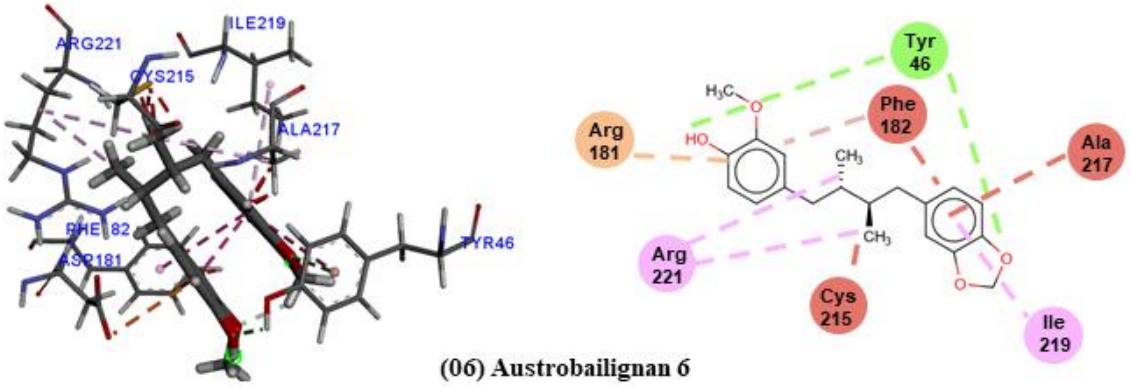
NQO1 must be activated to display antioxidant activity. According to Strandback et al. [127], the addition of N-(2-bromophenyl)pyrrolidine-1-sulfonamide (BPPSA) stabilized the flexible C-terminal region of the protein, resulting in the slower incorporation of deuterium. The amino acids involved in the bond were Tyr127, Thr128, and the catalytic residues Tyr156 and His162.

Compounds 6-[(2*R*,3*R*,4*R*,5*R*)-3,4-dimethyl-5-(3,4,5-trimethoxyphenyl)oxolan-2-yl] -4-methoxy-1, 3-benzodioxole (64) and Ococymosin (116) interacted well with PDE5. Compound 64 was able to form three hydrogen bonds with Met816, Tyr612 and Gln817, four hydrophobic interactions with the amino acids Cys677, Val782, Phe786 and Phe820. It also formed a steric interaction with Ile680. Ococymosin formed two hydrogen bonds with Tyr612 and Cys677 and five hydrophobic interactions with Ile680, Ala779, Val782, Phe786 and Phe820 (Figure 04).

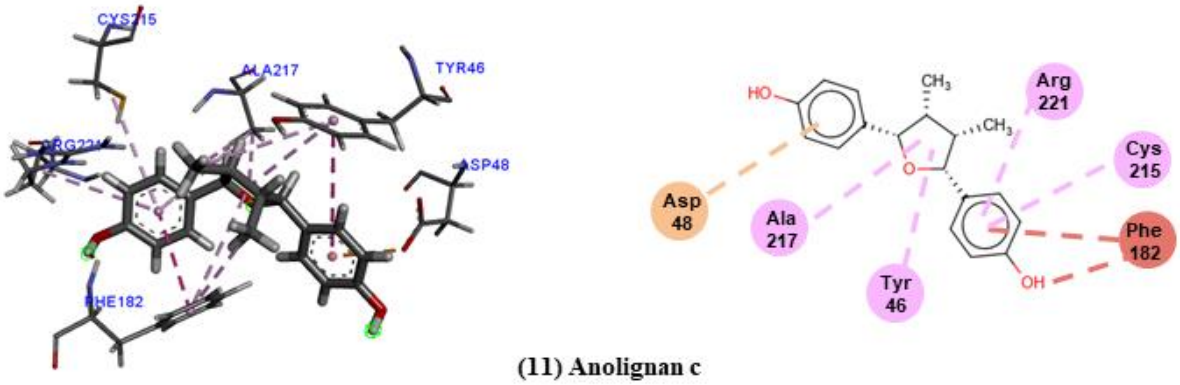
Experimental studies carried out by Wang et al. [128] showed that the drug vardenafil is a potent PDE5 inhibitor, binding to several amino acids in the active site. The amino acids that interacted with vardenafil are, Tyr612, Leu765, Ile768, Ala767, Ile680, Cys677, Ty676, Ile813, Met816, Gln817 and Phe820. Most of these amino acids also interacted with lignans.



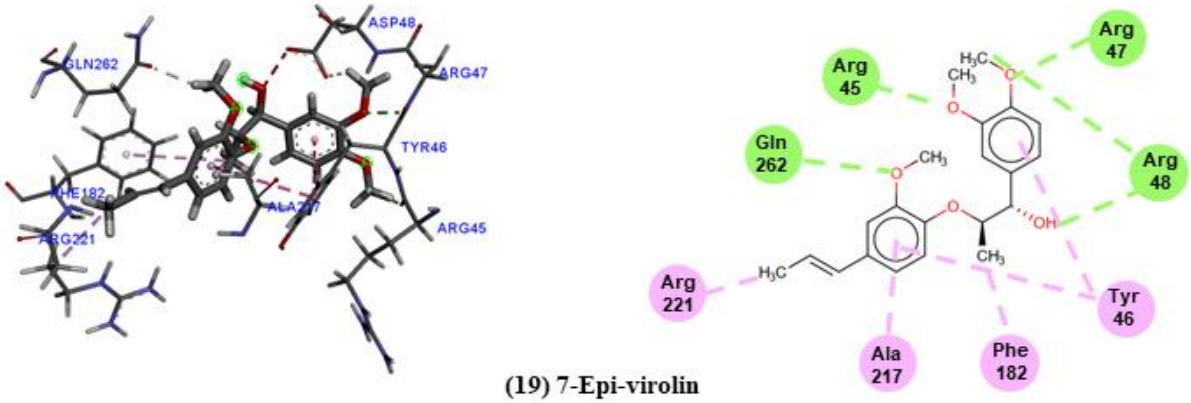
PTP1B



(06) Austrobailignan 6



(11) Anolignan c



(19) 7-Epi-virolin

Figure 2. 3D and 2D interactions between lignans and PTP1B. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, steric interactions are highlighted in red and electrostatic interactions are highlighted in orange.

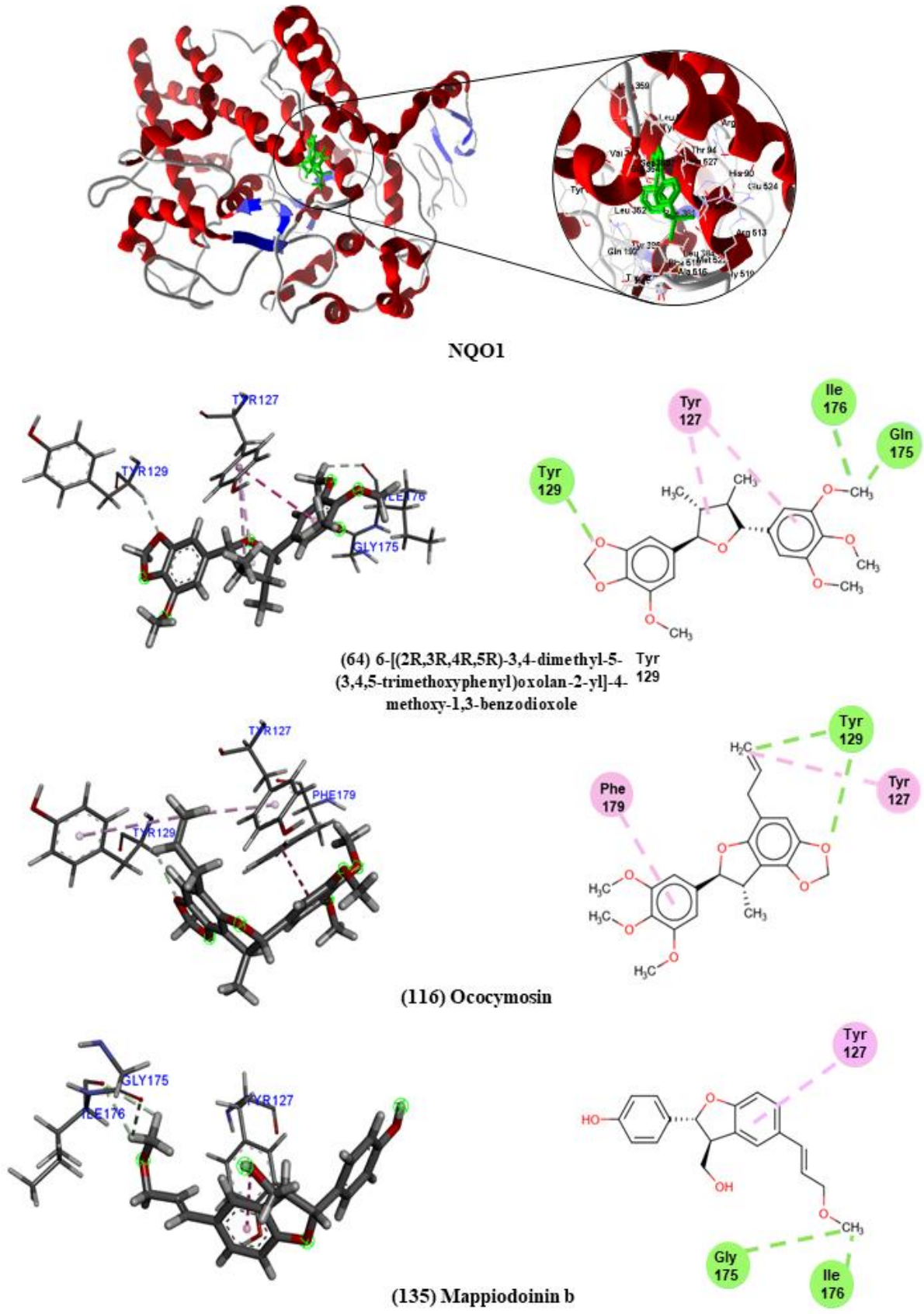


Figure 3. 3D and 2D interactions between lignans and NQO1. Hydrogen bonds are highlighted in green and hydrophobic interactions are highlighted in pink.

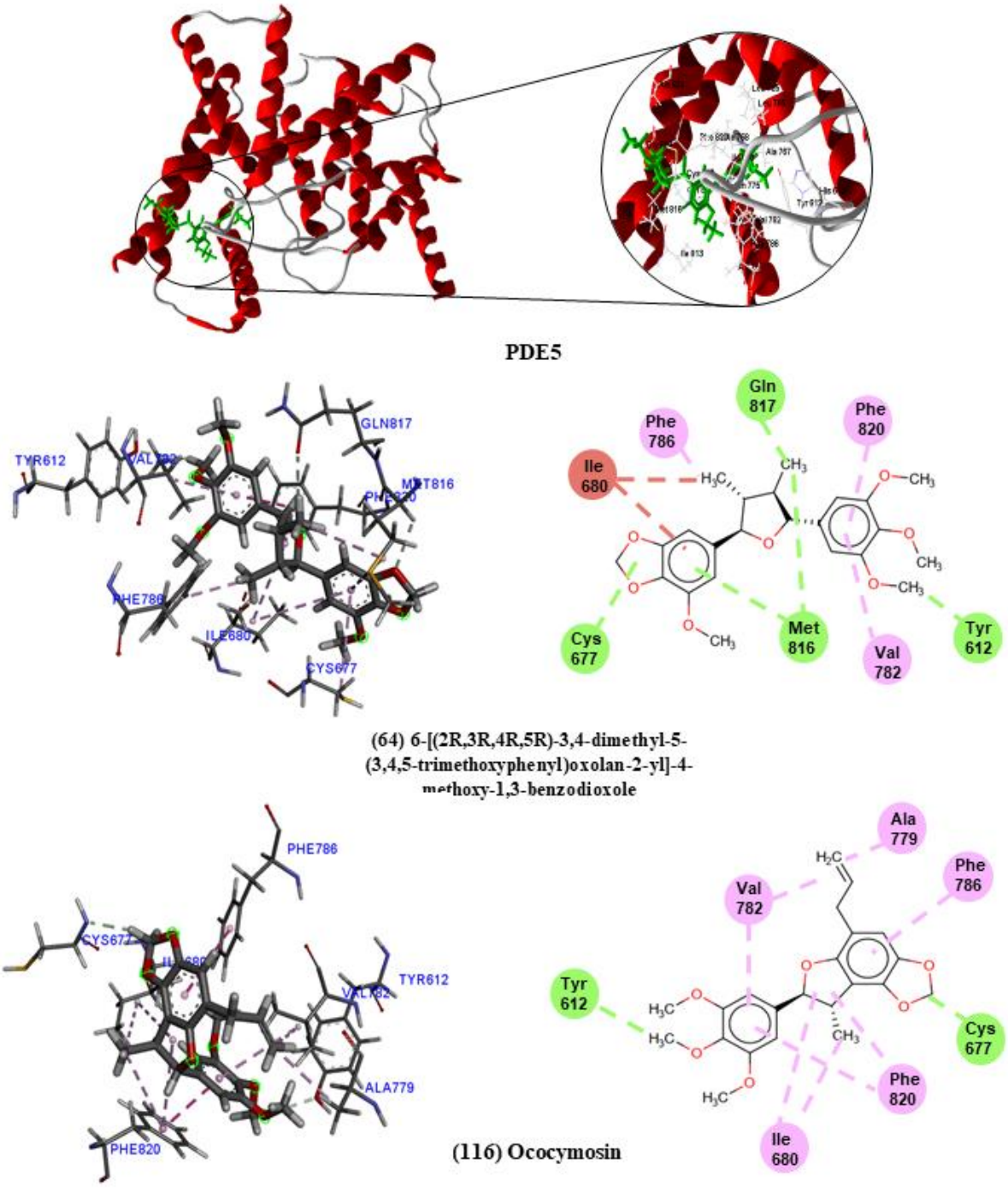


Figure 4. 3D and 2D interactions between lignans and PDE5. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink and steric interactions are highlighted in red.

4. CONCLUSIONS

AD is a complex and multifactorial disease, comprised of a variety of aberrant cellular and molecular processes in different cell types and brain regions. The activation and

inactivation of a variety of enzymes can contribute to neuroprotection or disease progression. Therefore, AD therapy must be able to block or compensate for various abnormal pathological events [38].

Few drugs are available for AD treatment. In addition, AD pathophysiology is not well-understood, and the identification of targets for disease treatment remains a major challenge for drug discovery. Therefore, in this study, we investigated several potential targets that are directly and indirectly involved in the development and progression of AD, through oxidative stress mechanisms, aiming to explore new targets and to design effective drugs, with minimal side effects, for AD treatment. We examined a set of lignans and used virtual screening to select compounds with potential multitargeting effects for the treatment of AD.

The predictive models built in this study obtained excellent performance results, with accuracies greater than 73%. To increase the predictive power and decrease the number of false positives generated by these models, a combined analysis was used, based on both ligand and structure. The combined analysis was able to identify potentially active molecules, based on the Random Forest and multitargeting models.

Out of 159 total lignans, several potentially active compounds were identified: three compounds with probabilities of activity ranging from 50%-61% for JNK-3, 43 compounds with a 52-72% probability for PTP1B, 57 compounds with 51%–72% probabilities for PDE5, 27 compounds with probabilities between 50%–61% for COX-2, and 27 compounds with probabilities between 50%–81% for *i*NOS, 111 compounds, with probabilities ranging from 50%–64%, were identified for Nrf2, nine compounds with probabilities ranging from 51%–78%, were identified for NOX1 and 156 compounds were selected, with probabilities ranging from 27%–100%, for NQO1. We also identified 139 potentially active molecules for two to five target enzymes, from the entire lignan set analyzed.

Among the 139 lignans that were considered to be potentially active and multitargeting, 92 showed good absorption, bioavailability, and solubility, ranging from soluble to moderately soluble. Among the compounds that were considered to be multitargeting we selected those likely to cross the blood-brain barrier, through an *in silico* evaluation, resulting in the identification of nine lignans, which were then evaluated for toxicity. The compounds austrobailignan (06), anolignan c (11), 7-epi-virolin (19), 6-[(2*R*,3*R*,4*R*,5*R*)-3,4-dimethyl-5-(3, 4,5-trimethoxyphenyl)oxolan-2-yl]-4-methoxy-1, 3-

benzodioxole (64), ococymosin (116), and mappiodoinin b (135) were considered to have no toxicity risks for the evaluated parameters.

We suggest that lignans, especially austrobailignan (06), anolignan c (11), 7-epi-virolin (19), 6-[(2*R*,3*R*,4*R*,5*R*)-3,4-dimethyl-5-(3,4,5-trimethoxyphenyl)oxolan-2-yl]-4-methoxy-1, 3-benzodioxole (64), ococymosin (116), and mappiodoinin b (135), have high probability of activity against several enzymes that may be involved in AD pathogenesis and may confer neuroprotective effects, with low toxicity. The proposed compounds are projected as possible solutions that needs to be validated experimentally.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors wish to acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Supplementary materials

Table S1: Smiles of the lignans used in the study.
Figure S1: ROC curves generated from the RF models, for each studied enzyme.
Table S2: The physical and chemical properties of lignans that are considered to be potentially active against Alzheimer's disease and have multitargeting effects against four or more enzymes.
Table S3: Pharmacokinetic properties of the lignans that are considered to be potentially active against Alzheimer's disease and have multitargeting effects against four or more enzymes. Lignans that are predicted to cross the blood-brain barrier are highlighted in bold.
Table S4: Toxicity evaluations for lignans with the best ADMET profiles, that are potentially active and that have multitarget effects against four or more enzymes. Lignans that did not present toxicity for any of the analyzed parameters are highlighted in bold.

REFERENCES

[1] M. Rosini, E. Simoni, R. Caporaso, and A. Minarini, “Multitarget strategies in Alzheimer’s disease: benefits and challenges on the road to therapeutics,” *Future Med. Chem.*, vol. 8, no. 6, pp. 697–711, 2016.

[2] R. Alvariño *et al.*, “Streptocyclinones A and B ameliorate Alzheimer’s disease pathological processes in vitro,” *Neuropharmacology*, vol. 141, no. September, pp. 283–295, 2018.

[3] S. Li, X. Zhao, P. Lazarovici, and W. Zheng, “Artemether Activation of AMPK/GSK3 β (ser9)/Nrf2 Signaling Confers Neuroprotection towards β- Amyloid-Induced Neurotoxicity in 3xTg Alzheimer’s Mouse Model,” *Oxid. Med. Cell. Longev.*, vol. 2019, 2019.

[4] F. Sivandzade, A. Bhalerao, and L. Cucullo, “Cerebrovascular and neurological disorders: protective role of NRF2,” *Int. J. Mol. Sci.*, vol. 20, no. 14, p. 3433, 2019.

[5] T. Suzuki and M. Yamamoto, “Molecular basis of the Keap1–Nrf2 system,” *Free Radic. Biol. Med.*, vol. 88, pp. 93–100, 2015.

[6] D. Zekry, T. Kay Epperson, and K. H. Krause, “A role for NOX NADPH oxidases in Alzheimer’s disease and other types of dementia?,” *IUBMB Life*, vol. 55, no. 6, pp. 307–313, 2003.

[7] N. Asiimwe, S. G. Yeo, M. S. Kim, J. Jung, and N. Y. Jeong, “Nitric oxide: Exploring the contextual link with Alzheimer’s disease,” *Oxid. Med. Cell. Longev.*, vol. 2016, 2016, doi: 10.1155/2016/7205747.

[8] D. Kumar, A. Ganeshpurkar, D. Kumar, G. Modi, S. K. Gupta, and S. K. Singh, “Secretase inhibitors for the treatment of Alzheimer’s disease: Long road ahead,” *Eur. J. Med. Chem.*, vol. 148, pp. 436–452, 2018.

[9] Y. Tang and W. Le, “Differential Roles of M1 and M2 Microglia in Neurodegenerative Diseases,” *Mol. Neurobiol.*, vol. 53, no. 2, pp. 1181–1194, 2016.

[10] N. Asiimwe, S. G. Yeo, M. S. Kim, J. Jung, and N. Y. Jeong, “Nitric oxide: Exploring the contextual link with Alzheimer’s disease,” *Oxid. Med. Cell. Longev.*, vol. 2016, 2016.

[11] M. T. Islam, “Oxidative stress and mitochondrial dysfunction-linked neurodegenerative disorders,” *Neurol. Res.*, vol. 39, no. 1, pp. 73–82, 2017.

[12] J. Chhetri, A. E. King, and N. Gueven, “Alzheimer’s Disease and NQO1: Is there a Link?,” *Curr. Alzheimer Res.*, vol. 15, no. 1, pp. 56–66, 2017.

[13] R. X. Santos *et al.*, “Nuclear and mitochondrial DNA oxidation in Alzheimer’s disease,” *Free Radic. Res.*, vol. 46, no. 4, pp. 565–576, 2012.

[14] A. Kumar and A. Singh, “A review on Alzheimer’s disease pathophysiology and its management: an update,” *Pharmacol. Reports*, vol. 67, no. 2, pp. 195–203, 2015.

[15] S. O. Bachurin, E. V Bovina, and A. A. Ustyugov, “Drugs in clinical trials for Alzheimer’s disease: the major trends,” *Med. Res. Rev.*, vol. 37, no. 5, pp. 1186–1225, 2017.

[16] J. Q. Kwong and J. D. Molkentin, “Physiological and pathological roles of the mitochondrial permeability transition pore in the heart,” *Cell Metab.*, vol. 21, no. 2, pp. 206–214, 2015.

[17] P. Albrecht, J. Lewerenz, S. Dittmer, R. Noack, P. Maher, and A. Methner, “Mechanisms of Oxidative Glutamate Toxicity: The Glutamate/Cystine Antiporter System xc¯ as a Neuroprotective Drug Target,” *CNS Neurol. Disord. - Drug Targets*, vol. 9, no. 3, pp. 373–382, 2012.

[18] T. H. Murphy, M. Miyamoto, A. Sastre, R. L. Schnaar, and J. T. Coyle, “Glutamate toxicity in a neuronal cell line involves inhibition of cystine transport leading to oxidative stress,” *Neuron*, vol. 2, no. 6, pp. 1547–1558, 1989.

[19] R. Briggs, S. P. Kennelly, and D. O’neill, “CMJv16n3-Briggs.indd,” *Clin. Med. (Northfield. Il).*, vol. 16, no. 3, pp. 247–53, 2016.

[20] M. C. Lu *et al.*, “Discovery of a potent kelch-like ECH-associated protein 1-nuclear factor erythroid 2-related factor 2 (Keap1-Nrf2) protein-protein interaction inhibitor with natural proline structure as a cytoprotective agent against acetaminophen-induced hepatotoxicity,” *J. Med. Chem.*, vol. 62, no. 14, pp. 6796–6813, 2019.

[21] A. Chaudhary, P. K. Maurya, B. S. Yadav, S. Singh, and A. Mani, “Current Therapeutic Targets for Alzheimer’s Disease,” *J. Biomed.*, vol. 3, pp. 74–84, 2018.

[22] A. García-Osta, M. Cuadrado-Tejedor, C. García-Barroso, J. Oyarzábal, and R. Franco, “Phosphodiesterases as therapeutic targets for Alzheimer’s disease,” *ACS Chem. Neurosci.*, vol. 3, no. 11, pp. 832–844, 2012.

[23] L. Liu, H. Xu, S. Ding, D. Wang, G. Song, and X. Huang, “Phosphodiesterase 5 inhibitors as novel agents for the treatment of Alzheimer’s disease,” *Brain Res. Bull.*, vol. 153, no. August, pp. 223–231, 2019.

[24] C. A. Peixoto, A. K. S. Nunes, and A. Garcia-Osta, “Phosphodiesterase-5 inhibitors: Action on the signaling pathways of neuroinflammation, neurodegeneration, and cognition,” *Mediators Inflamm.*, vol. 2015, 2015.

[25] A. F. Teich *et al.*, “PDE5 Exists in Human Neurons and is a Viable Therapeutic Target for Neurologic Disease,” *J. Alzheimer’s Dis.*, vol. 52, no. 1, pp. 295–302, 2016.

[26] S. Gourmaud *et al.*, “Increased levels of cerebrospinal fluid JNK3 associated with amyloid pathology: Links to cognitive decline,” *J. Psychiatry Neurosci.*, vol. 40, no. 3, pp. 151–161, 2015.

[27] X. Antoniou, M. Falconi, D. Di Marino, and T. Borsello, “JNK3 as a therapeutic target for neurodegenerative diseases,” *J. Alzheimer’s Dis.*, vol. 24, no. 4, pp. 633–642, 2011.

[28] R. Yarza, S. Vela, M. Solas, and M. J. Ramirez, “c-Jun N-terminal kinase (JNK) signaling as a therapeutic target for Alzheimer’s disease,” *Front. Pharmacol.*, vol. 6, no. JAN, pp. 1–12, 2016.

[29] A. Flemming, “Alzheimer’s Disease: JNK3 as new target in AD?,” *Nat. Rev. Drug Discov.*, vol. 11, no. 11, p. 829, 2012.

[30] H. Okazawa and S. Estus, “The JNK / c-Jun cascade and Alzheimer ’ s disease,” vol. 17, no. 2, pp. 79–88, 2002.

[31] Y. Wang *et al.*, “Curcumin reduces hippocampal neuron apoptosis and JNK-3 phosphorylation in rats with Aβ-induced Alzheimer's disease: protecting spatial learning and memory,” *J. Neurorestoratology*, vol. Volume 5, pp. 117–123, 2017.

[32] V. Tell *et al.*, “Drug Development of Small-Molecule Inhibitors of AD-Relevant Kinases as Novel Perspective Multitargeted Approach,” *Curr. Alzheimer Res.*, vol. 13, no. 12, pp. 1330–1336, 2016.

[33] J. Fiorito *et al.*, “Identification of a Novel 1,2,3,4-Tetrahydrobenzo[b][1,6]naphthyridine Analogue as a Potent Phosphodiesterase 5 Inhibitor with Improved Aqueous Solubility for the Treatment of Alzheimer’s Disease,” *J. Med. Chem.*, vol. 60, no. 21, pp. 8858–8875, 2017.

[34] J. Prickaerts, P. R. A. Heckman, and A. Blokland, “Investigational phosphodiesterase inhibitors in phase I and phase II clinical trials for Alzheimer’s disease,” *Expert Opin. Investig. Drugs*, vol. 26, no. 9, pp. 1033–1048, 2017.

[35] S. M. Nabavi *et al.*, “Phosphodiesterase inhibitors say NO to Alzheimer’s disease,” *Food Chem. Toxicol.*, vol. 134, no. September, p. 110822, 2019.

[36] W. Gulisano, M. R. Tropea, O. Arancio, A. Palmeri, and D. Puzzo, “Sub-efficacious doses of phosphodiesterase 4 and 5 inhibitors improve memory in a mouse model of Alzheimer’s disease,” *Neuropharmacology*, vol. 138, pp. 151–159, 2018.

[37] G. J. Song *et al.*, “A novel role for protein tyrosine phosphatase 1B as a positive regulator of neuroinflammation,” *J. Neuroinflammation*, vol. 13, no. 1, pp. 1–14, 2016.

[38] M. N. N. Vieira, N. M. Lyra e Silva, S. T. Ferreira, and F. G. De Felice, “Protein tyrosine phosphatase 1B (PTP1B): A potential target for Alzheimer’s therapy?,” *Front. Aging Neurosci.*, vol. 9, no. JAN, pp. 1–9, 2017.

[39] P. G. Través *et al.*, “Pivotal role of protein tyrosine phosphatase 1B (PTP1B) in the macrophage response to pro-inflammatory and anti-inflammatory challenge,” *Cell Death Dis.*, vol. 5, no. 3, pp. 1–13, 2014.

[40] K. M. Ricke *et al.*, “Neuronal Protein Tyrosine Phosphatase 1B Hastens Amyloid β-Associated Alzheimer’s Disease in Mice,” *J. Neurosci.*, vol. 40, no. 7, pp. 1581–1593, 2020.

[41] S. Sorce *et al.*, “NADPH oxidases as drug targets and biomarkers in neurodegenerative diseases: What is the evidence?,” *Free Radic. Biol. Med.*, vol. 112, no. August, pp. 387–396, 2017.

[42] M. W. Ma *et al.*, “NADPH oxidase in brain injury and neurodegenerative disorders,” *Mol. Neurodegener.*, vol. 12, no. 1, pp. 1–28, 2017.

[43] S. Shimohama *et al.*, “Activation of NADPH oxidase in Alzheimer’s disease brains,” *Biochem. Biophys. Res. Commun.*, vol. 273, no. 1, pp. 5–9, 2000.

[44] M. A. Ansari and S. W. Scheff, “NADPH-oxidase activation and cognition in Alzheimer disease progression,” *Free Radic. Biol. Med.*, vol. 51, no. 1, pp. 171–178, 2011.

[45] D. Ross and D. Siegel, “Functions of NQO1 in cellular protection and CoQ10 metabolism and its potential role as a redox sensitive molecular switch,” *Front. Physiol.*, vol. 8, no. AUG, pp. 1–10, 2017.

[46] A. K. Raina, D. J. Templeton, J. C. Deak, G. Perry, and M. A. Smith, “Quinone reductase (NQO1), a sensitive redox indicator, is increased in Alzheimer’s disease,” *Redox Rep.*, vol. 4, no. 1–2, pp. 23–27, 1999.

[47] K. S. SantaCruz, E. Yazlovitskaya, J. Collins, J. Johnson, and C. DeCarli, “Regional NAD(P)H:quinone oxidoreductase activity in Alzheimer’s disease,” *Neurobiol. Aging*, vol. 25, no. 1, pp. 63–69, 2004.

[48] P. Moi, K. Chan, I. Asunis, A. Cao, and Y. W. Kan, “Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region,” *Proc. Natl. Acad. Sci.*, vol. 91, no. 21, pp. 9926–9930, 1994.

[49] T. W. Kensler, N. Wakabayashi, and S. Biswal, “Cell Survival Responses to Environmental Stresses Via the Keap1-Nrf2-ARE Pathway,” *Annu. Rev. Pharmacol. Toxicol.*, vol. 47, no. 1, pp. 89–116, 2007.

[50] M. Abdalkader, R. Lampinen, K. M. Kanninen, T. M. Malm, and J. R. Liddell, “Targeting Nrf2 to suppress ferroptosis and mitochondrial dysfunction in neurodegeneration,” *Front. Neurosci.*, vol. 12, no. JUL, pp. 1–9, 2018.

[51] J. W. Kaspar, S. K. Niture, and A. K. Jaiswal, “Nrf2: INrf2 (Keap1) signaling in oxidative stress,” *Free Radic. Biol. Med.*, vol. 47, no. 9, pp. 1304–1309, 2009.

[52] D. A. Johnson and J. A. Johnson, “Nrf2 - A therapeutic target for the treatment of neurodegenerative diseases,” *Free Radic. Biol. Med.*, vol. 88, no. Part B, pp. 253–267, 2015.

[53] J. D. Hayes and A. T. Dinkova-Kostova, “The Nrf2 regulatory network provides an interface between redox and intermediary metabolism,” *Trends Biochem. Sci.*, vol. 39, no. 4, pp. 199–218, 2014.

[54] Y. Huang, W. Li, Z. Su, and A.-N. T. Kong, “The complexity of the Nrf2 pathway: beyond the antioxidant response,” *J. Nutr. Biochem.*, vol. 26, no. 12, pp. 1401–1413, 2015.

[55] E. M. Wright, D. D. F. L. LOO, and B. A. Hirayama, “Biology of human sodium glucose transporters,” *Physiol. Rev.*, vol. 91, no. 2, pp. 733–794, 2011.

[56] K. Shah, S. DeSilva, and T. Abbruscato, “The role of glucose transporters in brain disease: diabetes and Alzheimer’s disease,” *Int. J. Mol. Sci.*, vol. 13, no. 10, pp. 12629–12655, 2012.

[57] L. Szablewski, “Renal Glucose Transporters in Health and Renal Diseases,” *Biomed. Sci.*, vol. 24, p. 64, 2017.

[58] M. Sopjani, “The AMP activated protein kinase in the regulation of sodium coupled transporters (SGLT1, EAAT3 & EAAT4) and eryptosis.” 2010.

[59] Z. Li *et al.*, “Cardiac sodium-dependent glucose cotransporter 1 is a novel mediator of ischaemia/reperfusion injury,” *Cardiovasc. Res.*, vol. 115, no. 11, pp. 1646–1658, 2019.

[60] Y. J. Lee, J. S. Heo, H. N. Suh, M. Y. Lee, and H. J. Han, “Interleukin-6 stimulates α-MG uptake in renal proximal tubule cells: Involvement of STAT3, PI3K/Akt, MAPKs, and NF-κB,” *Am. J. Physiol. - Ren. Physiol.*, vol. 293, no. 4, pp. 1036–1046, 2007.

[61] Z. A. Jigheh *et al.*, “Empagliflozin alleviates renal inflammation and oxidative stress in streptozotocin-induced diabetic rats partly by repressing HMGB1-TLR4 receptor axis,” *Iran. J. Basic Med. Sci.*, vol. 22, no. 4, pp. 384–390, 2019.

[62] A. Mitsumoto and Y. Nakagawa, “DJ-1 is an indicator for endogenous reactive oxygen species elicited by endotoxin,” *Free Radic. Res.*, vol. 35, no. 6, pp. 885–893, 2001.

[63] D. Yu, H. Pan, R. Zhang, Y. Li, and X. Nie, “Nucleus DJ-1/Park7 acts as a favorable prognostic factor and involves mucin secretion in invasive breast carcinoma in Chinese population,” *Int. J. Clin. Exp. Med.*, vol. 10, no. 4, pp. 6558–6567, 2017.

[64] J. Fan, H. Yu, Y. Lv, and L. Yin, “Diagnostic and prognostic value of serum thioredoxin and DJ-1 in non-small cell lung carcinoma patients,” *Tumor Biol.*, vol. 37, no. 2, pp. 1949–1958, 2016.

[65] T. Taira, Y. Saito, T. Niki, S. M. M. Iguchi-Ariga, K. Takahashi, and H. Ariga, “DJ-1 has a role in antioxidative stress to prevent cell death,” *EMBO Rep.*, vol. 5, no. 2, pp. 213–218, 2004.

[66] M. Inden *et al.*, “Protection Against Dopaminergic Neurodegeneration in Parkinson’s Disease–Model Animals by a Modulator of the Oxidized Form of DJ-1, a Wild-type of Familial Parkinson’s Disease–Linked PARK7,” *J. Pharmacol. Sci.*, p. 1110280617, 2011.

[67] L. P. Dolgacheva, A. V. Berezhnov, E. I. Fedotova, V. P. Zinchenko, and A. Y. Abramov, “Role of DJ-1 in the mechanism of pathogenesis of Parkinson’s disease,” *J. Bioenerg. Biomembr.*, vol. 51, no. 3, pp. 175–188, 2019.

[68] R. Xue *et al.*, “DJ-1 activates autophagy in the repression of cardiac hypertrophy,” *Arch. Biochem. Biophys.*, vol. 633, pp. 124–132, 2017.

[69] C. R. del Mundo, A. L. Castillo, S. S. A. An, and M. A. Tan, “In vivo COX-2 modulation and metabolite profiling of Pandanus tectorius leaves extracts,” *3 Biotech*, vol. 10, no. 3, pp. 1–10, 2020.

[70] B. K. Baothman, J. Smith, L. J. Kay, S. K. Suvarna, and P. T. Peachell, “Prostaglandin D2 generation from human lung mast cells is catalysed exclusively by cyclooxygenase-1,” *Eur. J. Pharmacol.*, vol. 819, no. November 2017, pp. 225–232, 2018.

[71] M. Sandoughi, M. Saravani, M. Rokni, M. Nora, M. Mehrabani, and A. Dehghan, “Association between COX-2 and 15-PGDH polymorphisms and SLE susceptibility,” *Int. J. Rheum. Dis.*, no. December 2019, pp. 1–6, 2020.

[72] R. He *et al.*, “COX-2 mediated crosstalk between Wnt/β-catenin and the NF-κB signaling pathway during inflammatory responses induced by Haemophilus parasuis in PK-15 and NPTr cells,” *Dev. Comp. Immunol.*, vol. 105, no. 1, pp. 1–7, 2020.

[73] H. Wang, J. Wang, J. Wang, W. Cui, M. Sun, and Y. Liu, “Relationships of pain in pancreatic cancer patients with pathological stage and expressions of NF-κB and COX-2,” *J. B.U.ON.*, vol. 25, no. 1, pp. 448–453, 2020.

[74] A. Jorda *et al.*, “Action of low doses of aspirin in inflammation and oxidative stress induced by aβ1-42 on astrocytes in primary culture,” *Int. J. Med. Sci.*, vol. 17, no. 6, pp. 834–843, 2020.

[75] A. González-Barnadas, O. Camps-Font, P. Martín-Fatás, R. Figueiredo, C. Gay-Escoda, and E. Valmaseda-Castellón, “Efficacy and safety of selective COX-2 inhibitors for pain management after third molar removal: a meta-analysis of randomized clinical trials,” *Clin. Oral Investig.*, vol. 24, no. 1, pp. 79–96, 2020.

[76] E. M. Gedawy, A. E. Kassab, and A. M. El Kerdawy, “Design, synthesis and biological evaluation of novel pyrazole sulfonamide derivatives as dual COX-2/5-LOX inhibitors,” *Eur. J. Med. Chem.*, vol. 189, 2020.

[77] A. Buzharevski *et al.*, “Carboranyl Derivatives of Rofecoxib with Cytostatic Activity against Human Melanoma and Colon Cancer Cells,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.

[78] C. Brookes, W. J. Ribbans, L. Y. El Khoury, and S. M. Raleigh, “Variability within the human iNOS gene and Achilles tendon injuries: Evidence for a heterozygous advantage effect,” *J. Sci. Med. Sport*, vol. 23, no. 4, pp. 342–346, 2020.

[79] J. Li, H. Wang, J. Li, Y. Liu, and H. Ding, “LC-MS analysis of: Myrica rubra extract and its hypotensive effects via the inhibition of GLUT 1 and activation of the NO/Akt/eNOS signaling pathway,” *RSC Adv.*, vol. 10, no. 9, pp. 5371–5384, 2020.

[80] V. Wilmes *et al.*, “Increased inducible nitric oxide synthase (iNOS) expression in human myocardial infarction,” *Int. J. Legal Med.*, no. Mi, 2019.

[81] C. Chen *et al.*, “Limb ischemic preconditioning ameliorates renal microcirculation through activation of PI3K/Akt/eNOS signaling pathway after acute kidney injury,” *Eur. J. Med. Res.*, vol. 25, no. 1, pp. 1–9, 2020.

[82] M. O. Kseibati, G. S. G. Shehatou, M. H. Sharawy, A. E. Eladl, and H. A. Salem, “Nicorandil ameliorates bleomycin-induced pulmonary fibrosis in rats through modulating eNOS, iNOS, TXNIP and HIF-1α levels,” *Life Sci.*, vol. 246, no. December 2019, 2020.

[83] Y. Le *et al.*, “Liraglutide ameliorates palmitate-induced oxidative injury in islet microvascular endothelial cells through GLP-1 receptor/PKA and GTPCH1/eNOS signaling pathways,” *Peptides*, vol. 124, no. November 2019, 2020.

[84] N. Ahmad, M. Y. Ansari, and T. M. Haqqi, “Role of iNOS in osteoarthritis: Pathological and therapeutic aspects,” *J. Cell. Physiol.*, no. December 2019, pp. 1–11, 2020.

[85] S. S. Soskic, “Regulation of Inducible Nitric Oxide Synthase (iNOS) and its Potential Role in Insulin Resistance, Diabetes and Heart Failure,” *Open Cardiovasc. Med. J.*, vol. 5, no. 1, pp. 153–163, 2011.

[86] Q. Wang, H. Sun, X. Qi, and M. Zhou, “eNOS rs2070744 polymorphism might influence predisposition to hemorrhagic cerebral vascular diseases in East Asians: A meta-analysis,” *Brain Behav.*, no. January, pp. 1–8, 2020.

[87] P. Jiang, C. Li, Z. Xiang, and B. Jiao, “Tanshinone IIA reduces the risk of Alzheimer’s disease by inhibiting iNOS, MMP-2 and NF-κBp65 transcription and translation in the temporal lobes of rat models of Alzheimer’s disease,” *Mol. Med. Rep.*, vol. 10, no. 2, pp. 689–694, 2014.

[88] J. F. Gielis, L. Quirynen, J. J. Briedé, E. Roelant, P. Cos, and P. E. Y. Van Schil, “Pathogenetic role of endothelial nitric oxide synthase uncoupling during lung ischaemia-reperfusion injury,” *Eur. J. Cardio-thoracic Surg.*, vol. 52, no. 2, pp. 256–263, 2017.

[89] C. Schirra *et al.*, “Phosphorylation and activation of endothelial nitric oxide synthase by red fruit (Pandanus conoideus Lam) oil and its fractions,” *J. Ethnopharmacol.*, vol. 251, no. December 2019, pp. 3–10, 2020.

[90] S. A. Sonar and G. Lal, “The iNOS activity during an immune response controls the CNS pathology in experimental autoimmune encephalomyelitis,” *Front. Immunol.*, vol. 10, no. APR, 2019.

[91] X. Chen *et al.*, “Quercetin protects the vascular endothelium against iron overload damages via ROS/ADMA/DDAHII/eNOS/NO pathway,” *Eur. J. Pharmacol.*, vol. 868, no. December 2019, pp. 1–9, 2020.

[92] Z. Chen *et al.*, “CCL28-induced CCR10/eNOS interaction in angiogenesis and skin wound healing,” *FASEB J.*, no. August 2019, pp. 5838–5850, 2020.

[93] L. J. Zhu *et al.*, “Dentate nNOS accounts for stress-induced 5-HT1A receptor deficiency: Implication in anxiety behaviors,” *CNS Neurosci. Ther.*, vol. 26, no. 4, pp. 453–464, 2020.

[94] S. Choi, J. S. Won, S. L. Carroll, B. Annamalai, I. Singh, and A. K. Singh, “Pathology of nNOS-Expressing GABAergic Neurons in Mouse Model of Alzheimer’s Disease,” *Neuroscience*, vol. 384, pp. 41–53, 2018.

[95] J. Murciano-Calles, A. Coello, A. Cámara-Artigas, and J. C. Martinez, “PDZ/PDZ interaction between PSD-95 and nNOS neuronal proteins: A thermodynamic analysis of the PSD95-PDZ2/nNOS-PDZ interaction,” *J. Mol. Recognit.*, vol. 33, no. 4, pp. 1–11, 2020.

[96] G. A. Czapski, K. Czubowicz, J. B. Strosznajder, and R. P. Strosznajder, “The lipoxygenases: Their regulation and implication in alzheimer’s disease,” *Neurochem. Res.*, vol. 41, no. 1–2, pp. 243–257, 2016.

[97] M. Walther, I. Ivanov, G. Myagkova, and H. Kuhn, “Alterations of lipoxygenase specificity by targeted substrate modification and site-directed mutagenesis,” *Chem. Biol.*, vol. 8, no. 8, pp. 779–790, 2001.

[98] H. Kuhn and B. J. Thiele, “The diversity of the lipoxygenase family. Many sequence data but little information on biological significance,” *FEBS Lett.*, vol. 449, no. 1, pp. 7–11, 1999.

[99] Y. B. Joshi, P. F. Giannopoulos, and D. Praticò, “The 12/15-lipoxygenase as an emerging therapeutic target for Alzheimer’s disease,” *Trends Pharmacol. Sci.*, vol. 36, no. 3, pp. 181–186, 2015.

[100] I. Ivanov *et al.*, “Tight association of N-terminal and catalytic subunits of rabbit 12/15-lipoxygenase is important for protein stability and catalytic activity,” *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, vol. 1811, no. 12, pp. 1001–1010, 2011.

[101] J. Chu, P. F. Giannopoulos, C. Ceballos-Diaz, T. E. Golde, and D. Praticò, “5-Lipoxygenase gene transfer worsens memory, amyloid, and tau brain pathologies in a mouse model of alzheimer disease,” *Ann. Neurol.*, vol. 72, no. 3, pp. 442–454, 2012.

[102] Y. B. Joshi and D. Praticò, “The 5-lipoxygenase pathway: Oxidative and inflammatory contributions to the Alzheimer’s disease phenotype,” *Front. Cell. Neurosci.*, vol. 8, no. JAN, pp. 1–8, 2015.

[103] S. S. Karuppagounder *et al.*, "N-acetylcysteine targets 5 lipoxygenase-derived, toxic lipids and can synergize with prostaglandin E2 to inhibit ferroptosis and improve outcomes following hemorrhagic stroke in mice," *Ann. Neurol.*, vol. 84, no. 6, pp. 854–872, 2018.

[104] O. Firuzi, J. Zhuo, C. M. Chinnici, T. Wisniewski, and D. Praticò, "5-Lipoxygenase gene disruption reduces amyloid-β pathology in a mouse model of Alzheimer's disease," *FASEB J.*, vol. 22, no. 4, pp. 1169–1178, 2008.

[105] A. Di Meco, J. G. Li, and D. Praticò, "Dissecting the Role of 5-Lipoxygenase in the Homocysteine-Induced Alzheimer's Disease Pathology," *J. Alzheimer's Dis.*, vol. 62, no. 3, pp. 1337–1344, 2018.

[106] E. D. AlFadly *et al.*, "Tackling neuroinflammation and cholinergic deficit in Alzheimer's disease: Multi-target inhibitors of cholinesterases, cyclooxygenase-2 and 15-lipoxygenase," *Eur. J. Med. Chem.*, vol. 167, pp. 161–186, 2019.

[107] M. Estrada Valencia *et al.*, "Neurogenic and neuroprotective donepezil-flavonoid hybrids with sigma-1 affinity and inhibition of key enzymes in Alzheimer's disease," *Eur. J. Med. Chem.*, vol. 156, pp. 534–553, 2018, doi: 10.1016/j.ejmech.2018.07.026.

[108] D. Schweiger, G. Fürstenberger, and P. Krieg, "Inducible expression of 15-lipoxygenase-2 and 8-lipoxygenase inhibits cell growth via common signaling pathways," *J. Lipid Res.*, vol. 48, no. 3, pp. 553–564, 2007.

[109] A. Gaulton *et al.*, "ChEMBL: A large-scale bioactivity database for drug discovery," *Nucleic Acids Res.*, vol. 40, no. D1, pp. 1100–1107, 2012.

[110] D. Fourches, E. Muratov, and A. Tropsha, "Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research," *J. Chem. Inf. Model.*, vol. 50, no. 7, pp. 1189–1204.

[111] Dos Santos Maia M; de Sousa NF; Rodrigues GCS; Monteiro AFM; Scotti MT; Scotti L., "Lignans and Neolignans anti-tuberculosis identified by QSAR and Molecular Modeling," *Comb Chem High Throughput Screen.*, 2020.

[112] V. P. Lorenzo, J. M. B. Filho, L. Scotti, and M. T. Scotti, "Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B," *Brazilian J. Pharmacogn.*, vol. 25, no. 6, pp. 690–697, 2015.

[113] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "DRAGON software: An easy approach to molecular descriptor calculations," *Match*, vol. 56, no. 2, pp. 237–248, 2006.

[114] M. Hall *et al.*, "The WEKA Data Mining Software : An Update," vol. 11, no. 1, pp. 10–18.

[115] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme.," *Biochim. Biophys. Acta*, vol. 405, pp. 442–451, 1975.

[116] K. Roy, S. Kar, and P. Ambure, “On a simple approach for determining applicability domain of QSAR models,” *Chemom. Intell. Lab. Syst.*, vol. 145, pp. 22–29, 2015.

[117] W. F. J. Bitencourt-Ferreira, Gabriela; de Azevedo, “Molegro Virtual Docker for Docking.,” *Docking Screens Drug Discov.*, pp. 149–167, 2019.

[118] M. D. Brice, J. R. Rodgers, and O. Kennard, “The Protein Data Bank,” vol. 324, pp. 319–324, 1977.

[119] A. Daina, O. Michielin, and V. Zoete, “SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules,” *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–13, 2017.

[120] S. Mandal, M. Moudgil, and S. K. Mandal, “Rational drug design,” *Eur. J. Pharmacol.*, vol. 625, no. 1–3, pp. 90–100, 2009.

[121] V. M. Alves, R. C. Braga, E. N. Muratov, and C. Horta, “Quim. Nova,” vol. 41, no. 2, pp. 202–212, 2018.

[122] R. B. Teponno, S. Kusari, and M. Spiteller, *Recent advances in research on lignans and neolignans*, vol. 33, no. 9. 2016.

[123] F. Zálešák, D. J. Y. D. Bon, and J. Pospíšil, “Lignans and Neolignans: Plant secondary metabolites as a reservoir of biologically active substances,” *Pharmacol. Res.*, vol. 146, no. May, p. 104284, 2019.

[124] D. E. Jeffries and C. W. Lindsley, “Asymmetric Synthesis of Natural and Unnatural Dibenzylbutane Lignans from a Common Intermediate,” *J. Org. Chem.*, vol. 84, no. 9, pp. 5974–5979, 2019.

[125] E. Alfonzo and A. B. Beeler, “A sterically encumbered photoredox catalyst enables the unified synthesis of the classical lignan family of natural products,” *Chem. Sci.*, vol. 10, no. 33, pp. 7746–7754, 2019.

[126] N. Krishnan *et al.*, “PTP1B inhibition suggests a therapeutic strategy for Rett syndrome,” *J. Clin. Invest.*, vol. 125, no. 8, pp. 3163–3177, 2015.

[127] E. Strandback *et al.*, “A small molecule chaperone rescues the stability and activity of a cancer-associated variant of NAD(P)H:quinone oxidoreductase 1 in vitro,” *FEBS Lett.*, vol. 594, no. 3, pp. 424–438, 2020.

[128] H. Wang, M. Ye, H. Robinson, S. H. Francis, and H. Ke, “Conformational variations of both phosphodiesterase-5 and inhibitors provide the structural basis for the physiological effects of vardenafil and sildenafil,” *Mol. Pharmacol.*, vol. 73, no. 1, pp. 104–110, 2008.

Capítulo 6

O docking molecular é a principal técnica utilizada na triagem virtual baseada na estrutura. Diversas abordagens de docking consenso foram publicadas nos últimos 20 anos e foram desenvolvidas e empregadas para aumentar a confiabilidade dos resultados. Além disso, tamanha importância contribui para evitar que triagens virtuais de grandes bibliotecas virtuais de moléculas, que podem ser potenciais terapêuticos, sejam descartados durante as análises de docking molecular.

Por isso, neste capítulo, foram apresentados vários estudos que aplicaram o docking consenso sobre proteína-proteína, proteína-ligante, caracterização de interações e poses, rastreamentos paralelos de bibliotecas virtuais e combinações com modelos estatísticos e métodos de pontuação de consenso.

Consensus analyses in molecular docking studies applied to Medicinal Chemistry

Mayara dos Santos Maia[a], Gabriela Cristina Soares Rodrigues[a], Andreza Barbosa Silva Cavalcanti[a], Luciana Scotti[a], Marcus Tullius Scotti*[a]

[a]*Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, João Pessoa-PB, Brazil.*

Artigo publicado na revista: Mini Reviews in Medicinal Chemistry, 2020;20(14):1322-1340.

DOI: 10.2174/1389557520666200204121129.

Fator de impacto: 3.862

ABSTRACT: The increasing number of computational studies in medicinal chemistry involving molecular docking has put the technique forward as promising in the design of Computer Aided Drug Design. Considering the main method in the virtual screening based on the structure, consensus analysis of docking has been applied in several studies to overcome limitations of algorithms of different programs and mainly to increase the reliability of the results and reduce the number of false positives. However, some consensus scoring strategies are difficult to apply and in some cases are not reliable because of the small number of datasets tested. Thus, for such a methodology to be successful, it is necessary to understand why, when and how to use consensus docking. Therefore, the present study aims to present different approaches to docking consensus, applications and several scoring strategies that have been successful and can be applied in future studies.

Keywords: Molecular docking, consensus analysis, medicinal chemistry, virtual screening, consensus scoring strategies, statistical models.

1. INTRODUCTION

In virtual screening (VS) based on protein structure, molecular docking is the main technique capable of predicting the best tuning orientation of a ligand that binds to a target of interest to form a stable complex.[1, 2] Thus, in Medicinal Chemistry, this technique is widely used in the discovery of drugs, playing a fundamental role in the rational planning of drugs.[3] With the increasing number of publications related to molecular docking in the last

decades and the number of docking programs available, consensus docking approaches have been applied to increase the reliability of results and improve the hit rate in VS studies.

The consensus score consists of the combination of two or more docking analyses, contributing to decreasing the number of false positives and increasing the probability of true positives.[4] One of the early studies that used the consensus score was written by Charifson *et al*.[5] According to the authors, the combination of intersection-based punctuation functions provides a significant reduction in the number of false positives identified by individual scores, increasing the ability to identify active and inactive compounds.

Currently, the scoring methods used in docking are based on force field calculations, empirical functions or knowledge based on average force potential.[6], [7] All of these methods have their advantages and disadvantages, although they have been successfully employed in the design of Computer Aided Drug Design (CADD). However, several studies. [8]–[16] have shown that the combination of coupling results from different programs generates good performance in selecting the best poses and discrimination between active and inactive inhibitors when compared with individual docking.[17] According to the study by Houston *et al*.[9], hit rates in some docking programs ranged from 55 to 64%, while the consensus was able to predict 82% of the correct poses.

Problems related to protein and binder flexibility, conformational sampling and limitations of scoring functions of independent program configurations [2], [18] also encouraged researchers to use the consensus approach as a key point to overcome any deficiencies in VS. This is because the VS performance is highly dependent on the calculation settings, algorithm optimization and model selection, among others.[10] Combinations of docking results with quantitative structure–activity relationships (QSAR) methods have also been described in some studies to correlate docking scores and experimental affinities with linear regression analysis.[19], [20] This type of consensus method also contributes to increasing the probability of finding binders with therapeutic potential, reducing possible errors from coupling algorithms.

Although the consensus approach performs well, some consensus-matching methods may present obstacles: relatively small data set tested, some methods complicated and difficult to apply to new targets. [21] Thus, considering the best way to use the consensus

approach in studies of Medicinal Chemistry and VS, the present review aims to present different approaches, applications and strategies of consensus punctuation in molecular docking analyses that can be used in several kinds of research. Therefore, in this study, we initially sought to cite and describe several studies that applied docking consensus on protein–protein, protein–ligand, characterization of interactions and poses, parallel screenings of virtual libraries and combinations with statistical models. We then survey consensus scoring methods as the most relevant and widely used publications in scientific research.

2. DOCKING CONSENSUS APPLICATIONS

In recent years, several studies have reported the use of consensus docking to obtain results that are more reliable and overcome possible limitations of algorithms and scoring functions. Here, we describe a series of studies that used this method with scoring strategies.

2.1 Protein–protein consensus docking studies

Protein–protein consensus docking studies were performed by Spinello *et al*. [22] to verify the binding affinity of caspases that undergo dimerization by associating with the antagonist receptor interleukin-1 (IL-1Ra), thereby becoming inactive. The authors merged different search algorithms and punctuation functions to overcome errors in predicting protein–protein binding. Four web servers used for protein–protein docking were selected for the calculations: ZDOCK,[23] ClusPro,[24] pyDock[25] and HDOCK.[26] The RosettaDock web server[27] was used to refine docking postures obtained from the rigid coupling protocols. The results showed that the ZDOCK and PyDock servers predicted IL-1Ra binding to the caspase dimerization site in 57% of cases. This mode of connection was also observed with greater preference when using the PyDock, being responsible for 87% of the cases.

Kausar *et al*.[28] also investigated protein–protein interactions to analyze the residual contributions and reliability of the binding mode of the dynein 1-light chain protein complex (DYNLL1) and Pilin using consensus docking. To obtain the native conformation and to understand the mechanism of inflammatory action hypothetically triggered by the binding of the Pilin bacterial protein to human protein DYNLL1, in infections caused by *Pseudomonas aeruginosa*, AutoDock,[29] PatchDock,[30] FireDock,[31] ZDOCK[23] and

DOCK/PIERR[32] with refinement and rescoring of the algorithm FiberDock[33] and ClusPro[34] associated with molecular dynamics studies. According to the results, a new interaction was observed, revealing the involvement of the Lys130 residue in the complex binding and confirmation of stability due to structural rearrangements.

2.2 Protein–ligand consensus docking studies

Jaundoo and colleagues[34] used docking programs such as AutoDock[29], AutoDock Vina (Vina)[36] and Schrödinger's Glide (Glide)[37] to evaluate the immunological and hormonal potential of 43 drugs approved by the Food and Drug Administration (FDA) commonly used to treat the symptoms of Gulf War Disease (GWI). To remove outliers from the results, the median absolute mean deviation (MAMD) of each ligand pose obtained by the three docking programs of all crystal structures of a given target was first calculated. The MAMD formula[37]:

$$\text{MAMD} = \textit{median}\ (|\text{X}i\ \text{median} - (\text{X})|) \qquad (1)$$

where Xi refers to the free bond energy of the pose for the crystalline structure and X refers to the median of free bond energies of all programs of all crystal structures. The upper and lower limits were determined using the formula:

$$[\text{X}_{lower}, \text{X}_{upper}] = \textit{median}\ (\text{X}) \pm (3.5\ \text{MAMD}) \qquad (2)$$

where only extreme binding energies greater than a threshold of 3.5 times absolute deviations around the median were eliminated. Then, the Root Mean Square Deviation (RMSD) between the anchored and crystallographic poses was compared, using a cutoff score of 2.0 Å. According to the study, several drugs approved by the FDA were predicted as new hormone target ligands identified by consensus docking.

To understand better the interactions between CD2–CD58 proteins that are involved in various inflammatory and autoimmune processes, Leherte *et al.*[38] used computational methods to identify the electrostatic interactions that guarantee the specificity of binding between both proteins. Trajectories obtained from molecular dynamics simulations of CD58 complexes with three inhibitory peptides were used to characterize the effect of water molecules and hydrogen bonds on the positioning of each ligand on the CD58 surface. The docking consensus performed through the software AutoDock, Smina[39], Vina[10] and GOLD[40]; were used to characterize the interactions observed in experiments and in

molecular dynamics simulations. Although the docking results showed variability, it was possible to identify two important residues of the active site, Tyr86 in the ligand and Lys34 in the target, as already seen experimentally.

Another approach involving consensus docking and scoring selection strategies was the study by Olsson, García-Sosa and Ryde[41]. These researchers studied the binding affinity of 102 inhibitors to the farnesoid X receptor (FXR) using design data from Grand Challenge 2016 (GC2).[42] We initially estimated the mode and binding affinity for all 102 ligands with five different anchoring and punctuation methods. The programs used were: QPLD,[43] Glide SP,[36] Glide XP, AutoDock 4 (AD4) and Vina. After performing the docking, two different methods were used for the final score. In the first, a consensus score (SC) was used, involving the average of the five scores generated by the software. Second, a line with the classification of the averages of all scores (CR). Table 01 shows the docking results for each scoring function and the classification difference between SC and CR.

A different approach by Poli *et al*.[44] showed that FLAP (fingerprints for ligands and proteins) coupled to consensus docking may be a good alternative to identify potential inhibitors. The authors experimentally tested the reliability of this approach through the database of six active inhibitors for fatty acid amide hydrolase (FAAH) and processed in FLAP. The GRID molecular interaction fields (MIFs) calculated for each conformer and translated into fingerprints helped to construct a pharmacophoric model that served as a parameter for evaluating docking results and filtering commercial compounds. The 10 most promising compounds were purchased and tested and showed FAAH inhibition activity, achieving an 80% hit rate.

A study by Cui *et al*.[45] employed a supervised strategy to obtain a consensus of six scoring methods tested against a set of 77 resolved kinase structures. The functions used to obtain consensus score based on the binding power structure of the mTOR inhibitor (mammalian rapamycin target/signaling pathway regarded as a promising target for glioblastoma therapy) were: Chem-Score,[46] X-Score,[47] DrugScore,[48] DFIRE,[49] DOCK score[50] and Autodock score.[51] ChemScore and X-Score are knowledge-based potentials with statistical observations of close contacts between certain types of atoms or functional groups that occur more frequently than would be expected from a random distribution. DrugScore and DFIRE are empirical potentials that tell the number of various types of interactions in a protein-binding complex. The DOCK score and the Autodock score are

force field potentials describing protein-binding affinity by the sum of the van der Waals intermolecular force and electrostatic interactions between all the atoms of the protein and ligand. Autodock score, DOCK score, ChemScore and X-Score were calculated in the independent programs AutoDock, DOCK, GOLD and X-ScoreTM, respectively. DrugScore and DFIRE were calculated on the DSX-ONLINE and dDFIRE/DFIRE2 online servers, respectively. After the high throughput VS of potential mTOR ligands against various kinase inhibitors deposited in the kinase inhibitor protein database only 27 inhibitor compounds have been selected as promising candidates; these compounds are structurally diverse including azepine derivatives, imidazole, indole, pyridazine, pyrimidine, quinoline and the like. In particular, two flavonoids were identified as promising mTOR binders; the flavonoid compounds have a broad spectrum of kinase inhibitory activity.

2.3 Characterizations of interactions and ligands

Some consensus docking analyses are used only to confirm the best ligands of a virtual triage and characterize the main interactions with the active site residues. These forms of analysis generally use only two docking programs and do not require rigorous consensus scoring protocols.

Saikia *et al*.[52] analyzed the therapeutic potential of 177 phytochemicals against viral proteins of Influenza A subtype (H1N1). The software Mocker Virtual Docker (v.6.0)[53] and PLANTS[54] that use the Moldock score and PLP scoring functions, respectively, were selected for the docking studies. From the analysis, the method was able to select 12 compounds namely Octacosonal (135) for hemagglutinin (HA), acidic polymerase (PA) and nucleoprotein (NP); Cordifolioside A (128) for HA, ion channel (M2), neuraminidase (NA), PA and NP; Tinosporids (142) for HA, M2 and NP, Trisulfide, methyl-2-propenyl (70) for M 2, Nonacosane-15-one (134) for NA and nonstructural proteins (NEP/NS2), Rosmarinic acid (110), *cis*-sesquisabiene hydrate (160) and Vitexin (120) for NEP/NS2, Makisterone A (133) and Syringic acid for PA.

Another study performed by Sharma *et al*.[55] used two docking programs, Glide and GOLD, associated with other computational methodologies to select anti-tuberculosis compounds. The aim of the study was to identify the selective structure with the necessary characteristics to inhibit Dihydrofolate reductase from *Mycobacterium tuberculosis* (Mtb-DHFR) and to develop new compounds with selective inhibition of the target, through

computational studies and laboratory techniques. For this, a high throughput screening was performed initially using the Lipinski rule filter using a set of compounds from two virtual libraries, the National Cancer Institute (NCI) and the Drug Bank database. Of these, the 100 best results (NCI 72 and Drug Bank [56] were selected based on docking scoring, binding interactions, visual analysis and removal of duplicates and other molecules reported with Mtb-DHFR activity. The top 10 results were selected and analyzed visually for synthetic viability. Based on the synthetic viability hit-1 was selected, synthesized, directed and evaluated for biological activity.

Xing *et al*.[57] conducted research on the coagulation enzyme factor Xa (fXa) that plays a crucial role in the blood coagulation cascade. The authors used three-dimensional fragments-based drug design (FBDD) combined with the structure-based pharmacophore (SBP) model and structural consensus docking to identify new fXa inhibitors using two different docking programs: LigandFit in DS2.5[58] and Glide Schrödinger 2009. The molecules, which passed the ADMET properties and the pharmacophore screening, were prepared using the LigPrep module in Schrödinger 2009 and two parallel docking studies were performed independently to determine a consensual position for each molecule. The fXa framework used as the enzyme was obtained from the Protein Data Bank (PDB ID: 2W26). For each molecule, the RMSD value was calculated between the Glide and Ligandfit anchored poses. This value gives us a direct measure to evaluate the accuracy of fitting results. When the RMSD value between two fit results is greater than 3 Å, there is a high probability of being a false positive. Were analyzed 100 major compounds resulting from each docking program that were combined, and their poses were compared visually followed by calculation of the corresponding RMSD value. As a result, 43 compounds containing poses generated from the two different docking programs were considered valid docking results. After the VS, two 3780 and 319 hits with persistent high performance were identified. These compounds were synthesized as well as their analogues, and experimental data showed that most of the compounds exhibited potent *in vitro* activity against fXa.

Table 1. The top 10 results of the coupling with five different functions.

Ligand	XP	SP	AD4	Vina	QPLD	CS	CS rank	CR rank
1	-35.6	-32.6	-38.2	-36.8	-38.7	-36.4	53	70
2	-24.0	-33.3	-36.9	-37.7	-37.0	-33.8	71	79
3	-35.8	-33.7	-42.6	-45.2	-41.2	-39.7	46	50
4	-37.1	-26.8	-45.0	-39.7	-40.9	-37.9	50	54

5	-37.5	-31.7	-33.6	-35.1	-38.6	-35.3	59	73
6	-49.2	-43.7	-48.2	-50.2	-49.0	-48.1	14	13
7	-55.9	-49.0	-53.5	-55.6	-58.6	-54.5	1	1
8	-47.5	-42.6	-42.8	-51.5	-39.0	-44.7	29	35
9	-36.1	-41.2	-49.1	-54.8	-40.0	-44.2	30	33
10	-33.8	-33.6	-25.0	-26.8	-38.6	-31.6	80	92
11	-26.3	-31.6	-26.0	-22.2	-37.6	-28.7	89	99
12	-10.0	-35.7	-26.3	-24.7	-41.0	-27.5	92	83
13	-41.8	-45.4	-52.5	-58.6	-41.7	-48.0	15	11
14	-36.9	-42.8	-47.2	-53.1	-38.8	-43.8	35	40
15	-43.3	-26.8	-42.3	-38.9	-40.6	-38.4	48	52
16	-29.9	-26.5	-32.3	-29.3	-16.7	-26.9	95	95
17	-39.4	-37.2	-32.4	-24.7	-41.6	-35.1	63	59
18	-51.0	-36.9	-39.2	-44.8	-41.5	-42.7	41	38
19	-51.0	-42.9	-48.1	-51.5	-54.6	-49.6	10	12
20	-56.1	-44.2	-49.0	-50.2	-59.1	-51.7	5	6
21	-54.7	-44.0	-46.9	-55.2	-56.6	-51.5	6	5
22	-52.2	-41.6	-45.9	-50.2	-54.2	-48.8	11	14
23	-46.7	-38.5	-47.1	-42.3	-41.7	-43.3	39	30
24	-52.8	-46.7	-46.4	-57.3	-56.6	-52.0	4	3
25	-56.2	-41.3	-49.9	-56.5	-59.2	-52.6	2	2
26	-54.1	-39.3	-50.0	-53.6	-56.3	-50.7	9	9
27	-54.9	-52.0	-39.7	-51.5	-40.3	-47.7	18	18
28	-55.9	-50.2	-42.8	-54.4	-40.3	-48.7	12	10
29	-55.4	-48.6	-42.0	-53.6	-40.4	-48.0	16	15
30	-38.7	-45.4	-34.5	-44.8	-40.0	-40.7	45	49
31	-51.5	-48.3	-45.7	-49.4	-41.6	-47.3	19	15
32	-56.7	-43.0	-38.8	-40.2	-56.1	-47.0	21	20
33	-35.5	-16.7	-39.4	-23.8	-40.7	-31.2	82	71
34	-16.7	-17.4	-22.6	-25.9	-16.7	-19.9	101	101
35	-40.4	-47.1	-34.4	-43.5	-41.1	-41.3	44	43
36	-55.8	-53.8	-20.1	-42.7	-40.0	-42.5	42	41
37	-41.1	-41.0	-41.7	-43.5	-41.0	-41.7	43	42
38	-37.2	-23.8	-34.7	-26.8	-41.8	-32.8	75	62
39	-45.5	-43.8	-47.3	-54.0	-39.2	-46.0	24	23
40	-49.0	-40.8	-43.8	-53.1	-38.1	-45.0	28	37
41	-40.3	-35.4	-29.9	-23.4	-39.0	-33.6	73	77
42	-47.9	-40.3	-37.6	-53.1	-41.8	-44.1	31	26
43	-36.7	-23.2	-41.2	-37.2	-39.8	-35.6	58	60
44	-52.1	-41.6	-42.4	-41.8	-41.0	-43.8	33	29
45	-16.7	-16.7	-33.1	-15.9	-36.4	-23.8	98	100
46	-38.7	-37.7	-38.7	-23.8	-41.0	-36.0	55	56
47	-39.6	-36.6	-40.1	-31.4	-41.8	-37.9	51	48
48	-43.8	-36.7	-32.4	-25.1	-41.0	-35.8	57	57
49	-40.0	-35.8	-38.5	-21.8	-38.0	-34.8	64	72
50	-52.0	-41.5	-42.0	-48.1	-51.8	-47.1	20	21
51	-55.4	-39.9	-49.3	-52.3	-59.6	-51.3	7	8
52	-53.3	-49.2	-48.7	-51.5	-58.3	-52.2	3	4
53	-49.7	-50.3	-37.4	-52.3	-40.0	-45.9	25	27
54	-53.4	-47.3	-45.4	-54.0	-53.7	-50.8	8	7
55	-51.6	-43.1	-47.1	-51.0	-39.7	-46.5	22	22
56	-58.5	-41.5	41.4	-32.2	-62.9	-30.7	84	39
57	-47.4	-41.0	-44.6	-48.5	-16.7	-39.7	47	46
58	-57.6	-51.4	-37.9	-52.7	-16.7	-43.3	37	28

59	-48.2	-44.9	-29.2	-47.7	-16.7	-37.3	52	53
60	-56.5	-49.5	-42.8	-52.7	-39.5	-48.2	13	17
61	-43.6	-49.3	-40.9	-53.1	-38.5	-45.1	27	33
62	-43.5	-49.0	-39.4	-51.0	-56.2	-47.8	17	19
63	-44.1	-49.6	-33.5	-52.3	-40.8	-44.0	32	31
64	-49.9	-47.2	-34.4	-45.6	-39.2	-43.3	37	44
65	-9.2	-16.7	-42.2	-24.7	-16.7	-21.9	100	93
66	-50.7	-43.3	-37.9	-44.8	-40.7	-43.5	36	36
67	-42.2	-43.0	-11.8	-35.1	-41.0	-34.6	65	55
68	-49.9	-47.2	-33.2	-45.6	-39.2	-43.0	40	45
69	-40.9	-51.6	-15.8	-40.6	-41.3	-38.0	49	47
70	-52.0	-39.0	-49.5	-52.3	-39.7	-46.5	23	23
71	-43.9	-44.2	-39.2	-49.8	-41.8	-43.8	34	25
72	-55.4	-50.0	-36.9	-48.1	-39.2	-45.9	26	32
73	-34.0	-36.6	-36.3	-28.0	-40.5	-35.1	62	65
74	-38.1	-39.2	-26.7	-21.8	-39.0	-33.0	74	75
75	-24.8	-33.3	-35.8	-29.3	-37.8	-32.2	79	87
76	-39.2	-35.9	-28.4	-27.6	-41.2	-34.4	67	60
77	-32.5	-16.7	-24.5	-25.9	-38.9	-27.7	91	97
78	-34.9	-37.1	-20.3	-14.2	-36.4	-28.6	90	96
79	-21.7	-37.2	-29.5	-28.5	-16.7	-26.7	96	90
80	-41.4	-33.6	-31.1	-26.4	-16.7	-29.8	86	79
81	-37.0	-24.8	-26.0	-24.3	-39.7	-30.4	85	89
82	-40.9	-38.5	-23.0	-26.4	-41.8	-34.1	68	57
83	-27.3	-39.1	-27.2	-20.5	-39.7	-30.8	83	86
84	-36.2	-38.1	-25.6	-28.0	-40.9	-33.8	72	66
85	-14.6	-36.1	-29.9	-27.6	-40.4	-29.7	88	76
86	-16.7	-16.7	-13.2	-16.7	-33.6	-19.4	102	102
87	-34.6	-27.4	-29.7	-30.1	-39.4	-32.2	77	83
88	-14.5	-35.5	-24.8	-18.4	-38.5	-26.4	97	98
89	-40.0	-36.9	-23.6	-20.9	-41.6	-32.6	76	68
90	-38.7	-22.0	-42.8	-36.4	-41.8	-36.3	54	51
91	-35.8	-35.1	-39.0	-30.1	-39.7	-35.9	56	63
92	-33.7	-32.3	-35.8	-33.5	-37.6	-34.6	66	82
93	-33.3	-34.5	-36.9	-29.7	-34.9	-33.8	70	79
94	-32.7	-36.3	-39.2	-29.7	-38.4	-35.3	61	69
95	-33.4	-34.0	-19.5	-20.5	-41.7	-29.8	87	85
96	-42.0	-37.0	-20.5	-15.5	-41.8	-31.3	81	66
97	-32.5	-28.0	-31.1	-28.9	-16.7	-27.5	93	94
98	-40.3	-35.2	-22.3	-23.0	-40.2	-32.2	78	77
99	-36.0	-30.8	-37.0	-26.4	-39.4	-33.9	69	74
100	-38.7	-37.5	-38.0	-22.6	-39.6	-35.3	60	64
101	-22.1	-35.5	-36.3	-25.1	-16.7	-27.1	94	91
102	-45.0	-16.7	4.0	-11.7	-40.8	-22.0	99	87

(Source: Olsson, et al.,[436]).

A study performed by Park *et al.*[59] was about the α2 subunit of monophosphate-activated protein kinase (AMPK2) which serves as a promising target for the development of new drugs for stroke. The objective of this study was to identify a new class of AMPK2 inhibitors based on VS based on structure with docking and simulations. To complement

the shortcomings of the force-field and empirical scoring functions, a consensus scoring method was employed to select suspected inhibitors by the combined use of the AutoDock and FlexX programs.[60] A free-energy term of molecular solvation was increased for each scoring function. 260.000 library compounds were embedded simultaneously with the AutoDock and FlexX programs, in which 1000 compounds were selected with maximum score in each scoring function; of these a total of 118 compounds were included in the top virtual sets of both programs. Compounds included in both sets of virtual hits were then purchased from the compound supplier and tested for inhibitory activity against AMPK2 through enzyme inhibition assays. It was possible to identify seven structurally diverse AMPK2 inhibitors with micromolar inhibitory activity. Thus, the results of the present study demonstrate superior performance with the force field and empirical consensus scores that are modified to include the effects of binder solvation on docking.

Xing *et al*.[61] carried out a study to identify molecules with inhibitory activity against dipeptidyl peptidase IV (DPP-4), which is an important approach for the treatment of type 2 diabetes. The authors applied the screening method using two programs of docking: LigandFit in DS2.5[58] and Glide in Schrödinger 2009. DockScore and GScore were chosen to classify separately the poses embedded in each program. The high-level compounds of the last step were independently anchored in the DPP-4 structure by LigandFit and Glide. The 100 major compounds resulting from each anchoring program were combined, and their poses were visually compared followed by calculation of the corresponding RMSD. As a result, 51 compounds containing similar anchored poses were considered valid docking results. The presence of an important hydrogen interaction was verified for each compound as the hydrogen bond Glu205 and Glu206, playing a critical role in the interaction between inhibitors and DPP-4. Visual inspection of the interaction with Glu205 and Glu206 was defined as a prerequisite for potential DPP-4 inhibitors. Two successful compounds, HWL-405 and HWL-892, having high performance at all stages of VS were identified. These two successful compounds, along with several analogues were synthesized and evaluated for *in vitro* inhibition of DPP-4. Experimental results indicated that the most engineered compounds exhibited significant DPP-4 inhibitory activity.

Lodhi *et al*.[62] have conducted studies on the Rab3A protein, which is expressed predominantly in the brain and in synaptic vesicles. The anchoring of ligand molecules to the Rab3A structure using Glide was performed and cross-validation was done using

GOLD, with the aim of exploiting the Rab3A protein in the search for potential inhibitors. These programs have different search algorithms and punctuation functions. Thus, it was not possible to compare the fitness of GOLD and Glide directly. To base the estimates made by GOLD software, the X-Score consensus score program was used. [47] This scoring function computes a binding score for a given protein−ligand complex structure. The Glide and GOLD docking programs were used to select from the 1,364 molecules in the ZINC database with the best affinity for Rab3A. The five major molecules based on the Glide score ranking were selected. GTP docking was performed at the Rab3A binding site using the GOLD software. The Rab3A−GTP complex presented the best GOLD performance value of 77.73. For this, the Glide incremental construction algorithm and the genetic algorithm of GOLD were applied. The best-classified molecule (ZINC13152284) has a Glide score of −6.65 kcal/mol, X-Score of −3.02 kcal/mol and GOLD of 64.54 with three hydrogen bonds and nine hydrophobic contacts.

Further studies using two docking programs for consensus analysis and compound selection have been reported in the literature (Newcombe *et al*., 2018,[63] Aliebrahimi *et al*., 2018,[64] Onawole *et al*., 2017,[65] Hassaan *et al*., 2016,[66] Shah *et al*., 2015[67]).

3. PARALLEL DOCKING SCREENINGS

Consensus docking is a way to increase the quality of VS and when applied to compounds from large virtual libraries, the approach based on parallel screenings may contribute to the search for molecules with therapeutic potential. Here, the term “parallel triages” was used to refer to the use of different docking programs in simultaneous triages to select the best compounds at each triage. Thus, by using more than one type of docking program, this approach can be considered a form of consensus analysis, such as the VS performed by Onawole *et al*. [4]. In this study, a set of 35,742,734 compounds from the Mcule database[68] was initially selected for its absorption, distribution, metabolism, excretion and toxicity (ADMET) properties. Then, they were subjected to a VS using Vina,[35] obtaining 981 incoming ligators for the second VS with FlexX,[60] both against *Zaire ebolavirus* (Ebola) glycoprotein. A voting classification approach[69] was used for the consensus score that compared the top 25 labeled ligands in both anchor protocols. The study considered the compound SC-2 as the most remarkable with properties regarding drug similarity, metrics of binder efficiency, solubility, absorption and distribution properties and non-carcinogenicity as observed in Fig. (**1**).

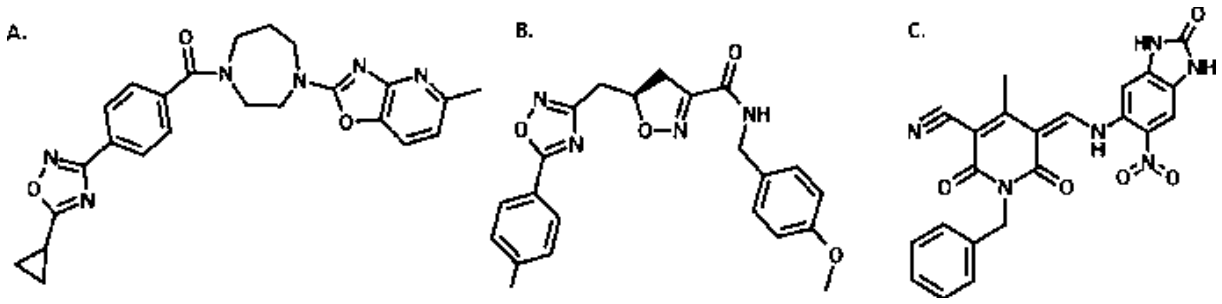


Fig. (1). Structures of SC-1 (A) SC-2 and (B) SC-3 (C) ligands selected from the consensus score of FlexX and AutoDock Vina (Source: Onawole *et al*. [3]).

Another article that also applied parallel screenings with protocols of consensus analysis of docking; involving a database of natural products was produced by Gangopadhyay, Chakraborty and Datta.[70] The aim of the study was to identify serotype variations in β-OG drug susceptibility due to dissimilarities of serotype sequences against dengue and to propose potential specific compounds. A set of compounds with 9,175 alkaloids with good ADMET properties were selected for the three steps of docking. In the first step, the alkaloids were coupled against four serotypes of β-OG bags from four dengue serotypes using Glide SP.[36] In the second step, 10% of the best results from the previous step were cross-docked with Glide XP.[36] Likewise, 10% of the best results and with RMSD ≤ 2 Å from the previous step were selected for the third screening using the AutoDock v4.2.[29] The consistency of the interaction was measured in terms of hydrogen bond identity and similarity of hydrophobic interactions. The best poses with the best interactions were selected for experimental tests. The results and screening protocols can be seen in Table (**2**).

Mamidi, Arora and Surolia[71] used several computational resources to select compounds against *Plasmodium falciparum* (PfFabD) and *Toxoplasma gondii* (TgFabD) FabD proteins and to map the chemical interaction space. VS was performed initially with Autodock 4.2.[29] The best results served as input for the second virtual screening with DOCK 6.[72]The best conformations of the AUTODOCK calculations validated by the DOCK 6 were recovered based on the low bond energies and classified using X-Score.[47] The study identified 160 and 3288 compounds for PfFabD and TgFabD, respectively. Further, a third filter was used based on pharmacokinetic properties. Finally, a set of 60 ligands for PfFabD and 131 ligands for TgFabD passed the ADMET filters.

Huang *et al*.[73] evaluated the performance of four docking programs using a set of molecules from two virtual libraries, Specs and ChemDiv, totaling 550.0 molecules. Initially, the compounds were submitted to molecular docking through the LibDock, GOLD, CDOCKER, and LigandFit software, incorporated in the Discovery Studio (DS) 3.1 platform (Accelrys Inc., San Diego, CA, USA) against structures of SIRT2. The authors observed that among all four coupling programs, GOLD and CDOCKER exhibited lower RMSD values for all binders studied. Thus, GOLD and CDOCKER were selected for the second stage of docking with the 22 compounds obtained in the first screening with good IC_{50} bioactivity. Four types of GOLD and CDOCKER scoring functions were used in docking analyses. The Spearman correlation coefficient (Rs) was used as a metric to measure the performance of punctuation functions, correlating the bioactivity and punctuation values of the compounds. The results showed that the score function with the lowest Rs was ChemScore. In addition, five compounds (SR13, SR17, SR10, SR30 and SR40) showed an inhibition rate >30% against SIRT2 at the concentration of 10 μM. The SR17 compound was considered the most active and supported structure and activity–structure optimization (SAR) that allowed the discovery of a new active compound, the triazino [5,6-b] indole.

Piperin is a phenolic compound with several activities described in the literature; among them, antioxidant, antitumor, anti-inflammatory, antidiarrheal, among others. To identify new anticancer compounds; Sattarinezhad, Bordbar and Fani[74] evaluated the inhibitory potential of piperin derivatives against Survinine. Considered a marker of cancer and a promising therapeutic target, Survinine is one of the proteins responsible for apoptosis. Virtual trials based on consensus docking, molecular dynamics simulations and free-energy calculations were used to select the best binders. The predicted binding energy results ranked the major compounds using the AutoDock Vina. Then a second screening using the AutoDock program ranked the 100 best molecules, in which 20 of these were with RMSD values low and between the top 100. Two molecules out of 20 were submitted to the simulations and calculations of free energy and confirmed their inhibitory potential higher. To find potent inhibitors with excellent ADMET properties against the enzyme β-secretase 1 (BACE1); Ju *et al*.[75] developed a detailed methodological protocol. They considered a target located in the brain and indicated as a target in Alzheimer's treatment, pharmacophores were generated based on VS. Generated pharmacophoric models were validated by comparing IC_{50} values with adjustment values of compounds in the test

compound pool. The optimized model was applied for the Chemdiv and Zinc database tracing. Compounds that underwent pharmacophoric screening were submitted to the LigandFit and CDOCKER programs of Discovery Studio 3.1.

Table 2. Virtual screening workflow and the number of compounds screened in each step against serotype specific β-OG.

				SCREENED ALKALOID LIBRARY (9175 ALKALOID)			
PHASE	**SCREENING PROTOCOL**	**SCREENING CRITERIA**	**HIT/LEAD**	**NUMBER OF HITS/LEADS AGAINST SEROTYPE**			
				1	**2**	**3**	**4**
I	Glide SP algorithm, Glide SP scoring function	Top ~10% scoring hits	Phase I virtual hits	894	903	906	958
II	Glide XP algorithm, Glide XP scoring function	Top ~10% scoring hits Glide SP-XP pose RMSD ≤2.0 Å	Phase II virtual hits	87	95	92	99
III	Lamarckian genetic, algorithm (LGA), AutoDock scoring function	Top ~10% scoring hits Glide (LGA) pose RMSD ≤2.0 Å	Phase II virtual hits	17	15	42	12
IV	Ligand-receptor interaction consistency	Identical hydrogen bonds, similarity of hydrophobic interactions	Phase III virtual hits	4	2	6	2
V	Comparison with experimental standards	Binding energy cut-off	Virtual leads	2	1	4	1

(Source: Gangopadhyay, Chakraborty e Datta [59]).

Compounds with a consensus score ≥1 and good interaction patterns were chosen for ADMET properties analysis. Of the three compounds with high LogP, only one showed in vitro inhibitory potential against BACE1.

Other studies using simultaneous screening as consensus analysis have also been described in the literature (Berry; Fielding; Gamieldien, 2015[76] and Wang *et al*., 2015[77]).

4. SCORING CONSENSUS

Some computational studies use consensus scoring as a method of selecting the best compounds in a VS. In this review, the use of different scores from the same software as a consensus approach has been described.

Although Pini *et al.*[78] used only two docking programs, GOLD 5.1 and PLANTS[54], four of the five scoring functions used in the study belong to GOLD. These are ASP, CSCORE, CHEMPLP and GSCORE. In addition, the CHEMPLP scoring function is very similar to the PLANTS scoring function PLP. Therefore, in this review, this study was applied to the consensus scoring approach. According to the authors, each binder from a series of six synthetic compounds was docked to the binding site against the enzyme

salicylate synthase of *Mycobacterium tuberculosis* (MbtI) using the different docking procedures. Then, the RMSD of each of these berthing positions against the remaining docking results was evaluated using the GOLD suite's RMS analysis program. The best pose in common for each coupling was considered the best docking pose. Compound 1 (Fig. **2**) obtained the best *in vitro* results and common interactions were observed compared with the reference inhibitor.

Dutta *et al*. [79] combined seven scoring functions (GoldScore, ChemScore, Aspscore, ChemPLP, Glidescore, Autodock score and FlexX score) of four docking programs: GOLD, AutoDock, Glide and FlexX. In addition, they used several consensus scoring strategies to select the best results. The main objective of this study was to understand polypharmacological profiles of drugs against kinases by establishing a computational pipeline using several consensus scoring strategies for the identification of therapeutic molecules. A set of 233 kinases was used for the reverse sorting method for the target profile. To normalize the scores of the moored complexes, a method was applied in which the absolute scores obtained from different punctuation functions for each embedded complex were auto-scaled between 0 and 1 using the formula:

$$X_{scaled} = (X_i - X_{min})/(X_{max} - X_{min}) \qquad (3)$$

Here, X_{scaled} is the normalized score *i*; X_i is the gross value of the *i* score; X_{min} is the smallest and X_{max} the highest value of 1080 anchored models. The consensus-ranking strategies selected were referred to as rank-by-number, rank-by-rank and rank-by-vote.[6], [80] The authors concluded that consensus-based scoring strategies improve the sensitivity of the reverse sorting approach, ensuring an optimal hit rate with high accuracy. With this, it was possible to find the true kinase targets for 12 known drugs.strategies selected were referred to as rank-by-number, rank-by-rank and rank-by-vote.[6], [80] The authors concluded that consensus-based scoring strategies improve the sensitivity of the reverse sorting approach, ensuring an optimal hit rate with high accuracy. With this, it was possible to find the true kinase targets for 12 known drugs.

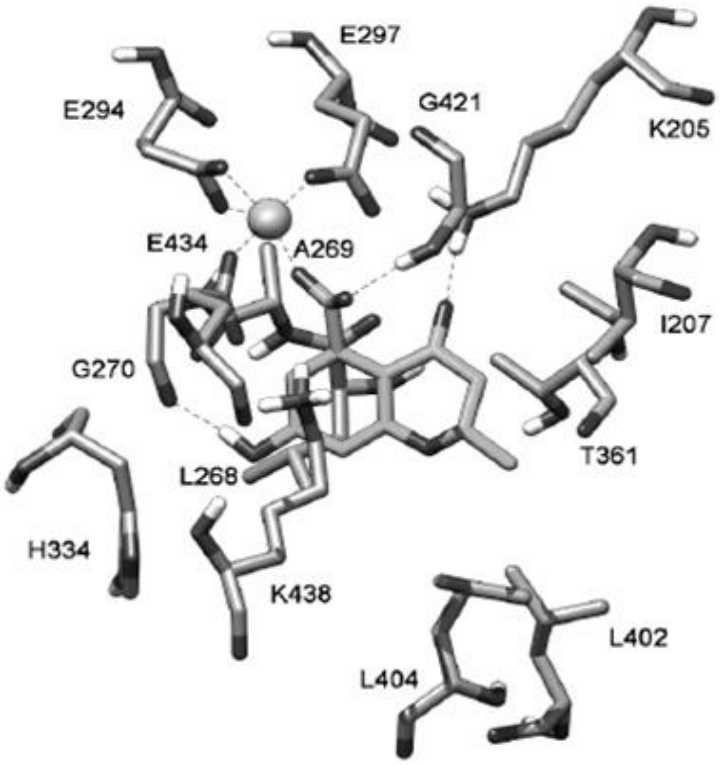


Fig. (2). Minimized mean structure of compound 1 anchored at the MbtI binding site (Source: Pini *et al*.[78]).

Raj *et al*.[81] in their study proposed the inactive human rhomboid protease 2 (IRHOM2) structure, which is a promising target in the discovery of new therapeutic drugs to cure cancer and rheumatoid arthritis. The authors subjected the refined modeling structure to molecular dynamics simulation and docking analysis. The VS were performed using Glide[36] with libraries containing 24552 compounds and the analysis presented a large network of hydrogen bonds and hydrophobic interactions that play an important role in their binding. The docking results were analyzed for high-rank compounds using a consensus-based score to calculate binding affinity as a measure of protein–ligand interactions, resulting in 10 potent inhibitors. Looking at the ligand binding results shows that amino acid residues Lys 358, Arg 384, Thr 409, Glu 580, Phe 662, Lys 638, Trp 822, Tyr 826 are critical for protein–ligand binding. The 10 new inhibitors chosen have been found to have much better scores and interactions than known protease inhibitors, showing potential for inhibition probability. The main molecule against the active site of IRHOM2 has a g-score of 12.565 kcal/mol and a glide e-model score of –74.967 kcal/mol with three hydrogen bonds and 11 hydrophobic interactions. They therefore, concluded that this molecule may

act as a likely inhibitor against these chronic diseases, but further studies are needed *in vitro*.

In the work developed by Fourches *et al*.[82], a single consensus prediction approach was employed in two ways: first, molecular docking predicted the binding positions of the binders and CSAR 2013 ranked them according to their docking scores that were constructed using chemical descriptors of the protein-binding interface. To carry out the CSAR 2013 benchmark exercise, the authors applied a hybrid docking and punctuation workflow to classify 10 steroid binders of a modified digoxigenin-binding protein. The Glide software was used to provide poses for each steroid binder and to classify based on standard precision (SP) docking and extra precision (XP) docking scoring functions. The poses were encoded by the MCT-Tess descriptors of the protein-binding interface, and by random forest (RF) models trained to distinguish the two classes of poses based on their descriptors. The consensus classifier was added to the Glide-generated poses of each CSAR binder, in filter the predicted poses and recreate the remainder using the XP and SP scoring functions. For the final classification of the ligand, the best punctuation pose was used for each ligand after this selection step. The accuracy of the ranking for the 10 binders evaluated by the Spearman correlation coefficient was 0.64 for SP and 0.52 for XP but reached 0.75 for SP/RF consensus score (ranked third in the CSAR 2013 reference). Through this study, it was possible to verify that the specific goal-scoring models are able to increase the reliability of structure-based molecular docking, discarding the false poses.

5. STATISTICAL METHODS IN PUNCTUATION CONSENSUS

Some studies have reported the use of principal component analysis (PCA), partial least squares (PLS) regression, Bayesian classification and machine learning in molecular docking consensus. These models contribute to the analySes because the scores generated by the mooring are quite crude and therefore not very effective in the fine classification of active compounds based on their affinities or powers.[19]

They are generally suitable for distinguishing active compounds from inactive. While the models have the ability to distinguish between active and inactive compounds, in addition to indicating the degree of potency of each compound. The receiver operating characteristic (ROC) curves obtained in these pilot trials can be

conveniently used to optimize the coupling protocol and select the most appropriate punctuation functions for a given target. [19]

An example is the study by Li *et al*.[83], who constructed models based on a series of consensus scores through a set of scoring functions. A set of 25 mammalian rapamycin target inhibitors (mTOR) were selected for the molecular fitting study using the Glide program.[36] The predicted postures of these ligands were saved and re-evaluated by 21 punctuation functions available in various programs. To improve the quality of the models, three types of binder efficiency indexes (BEI, SEI and LLE) were applied instead of only pIC_{50} data:

$$BEI = pActivity/NA \quad (4)$$

where NA is the number of non-hydrogen atoms within the molecules while the pActivity value is calculated in the same way as pIC50.

$$LLE = pActivity - ALOGP \quad (5)$$

where ALOGP, the important part for LLE, is calculated by the Canvas module of the Schrodinger 2015 suite.

$$SEI = 100\% \times pActivity/TPSA \quad (6)$$

where TPSA, is the key property of SEI.

Next, punctuation functions were used to perform a QSAR Contingency calculation on the MOE2014 for the selection of descriptors that make up the punctuation functions. Subsequently, the consensus scoring models were trained based on the results obtained by the (PLS) method, and validated by Leave-one-out (LOO).

The two best models calculated were 10 and 22 (Eq. 7 and 8, respectively):

$$BEI = 0.70099 - 0.05462 \times LigScore2 + 0.00002 \times LUDI\ 1 + 0.00115 \times DScore \quad (7).$$

(Training: $R^2 = 0.767$, $Q^2 = 0.647$, RMSE = 0.024, e test: $R^2 = 0.932$, RMSE = 0.026).

$$BEI = 0.71093 + 0.00050 \times Alpha\ HB + 0.00072 \times DScore\ 0.03113 \times LigScore2 - 0.01201 \times XScore \quad (8).$$

(Training: R^2 = 0.790, Q^2 = 0.627, RMSE = 0.023, e test: R^2 = 0.955, RMSE = 0.020).

Alzweiri *et al*. [84] used an HPLC size-exclusion method which was prepared as an assay procedure to analyze the binding of compounds tested with carbonic anhydrase III (CAIII) enzyme. The consensus scoring strategy was performed in this study, using different docking programs, namely, CDOCKER, LibDock, LigandFit and GOLD, because each uses the pose in different methods of adaptation of the ligand within the active site with its own energy calculation methodology. Thus, using such techniques to minimize any tendency that may arise in a single coupling program, the affinity values of 20 training compounds were modeled against 35 descriptors derived from their structures. The external validation of the model was tested by the analysis of six molecules through the chromatographic method and compared with the corresponding predicted values in the QSAR model. Twenty-six compounds were used between the training and test series. Inhibition of CAIII by a group of benzoic acid analogues was characterized by the (negative) peak intensity vacancy representing the fraction of the compounds bound to the CAIII enzyme. Interestingly, *p*-hydroxylbenzoic acid and aspirin were found to be potent inhibitors against CAIII. Docking studies in the compound training set generated consensus scores with strong affinity factors obtained from the chromatographic analysis, as well as also showed the importance of the negative charged carboxylates.

Ericksen *et al*. [85] compared common consensus scoring methods with a new form of unsupervised gradient reinforcement. They also observed a greater difference in score between active ligands and created a consensus statistical model score based on the combination of mean scores and variances. They used the common performance metrics ROCAUC and EF1 in 21 DUD-E reference targets for performance evaluation. To develop two consensus scores from eight distinct docking scores, we used blended modeling and gradient enhancement. Entries for the model are not all link energy predictions because the scores are based on a variety of scoring approaches. Therefore, the consensus scores provided are also not predictive of binding energy, but they do form qualifying scores to enrich a small selection with active compounds from a large library. In relation to the usual methods of consensus, the mixing model and the increase of the gradient provide additional improvements. Therefore, they are considered methods that can be applied in new targets in academic research and are considered efficient when compared with the use of a single method of docking in a new target.

Cotesta *et al*.[86] performed an evaluation of two docking programs (QXP and GOLD [435]), and compared the ability of six simple punctuation functions (PLP, Ligscore, Ludi, Jain, ChemScore and PMF) and four scoring models (Mean Rank: MR, Rank-by-Vote: Vt, Bayesian Statistics: BS and PLS Discriminant Analysis: DA) and thus succeeded in separating compounds that are active against inactive CDK2. The objectives of this work were to verify that a good prediction of affinity requires a good prediction of pose by comparing and analyzing score results based on poses generated by QXP and GOLD, and also to determine for each program and range of activities, which scoring schemes produce the best results to differentiate the active compounds from inactive. It was possible to determine the enrichment for the entire set of assets (IC_{50} <10M) and for three subsets of activities. As a result, it was found that the enrichment for each subset was smaller than for the whole set of assets. QXP is better than GOLD in the pose forecast and with the settings that were used, GOLD was much faster than QXP. It was found that for each program, two scoring functions generally performed better than the others (Ligscore2 and Ludi for GOLD, QXP and Jain for QXP) and that composite punctuation functions produced better results than simple punctuation functions. The consensus is closer to MR and Vt which were better at separating the micromolar inhibitors from the inactive ones. Statistical approaches (BS and DA), which require training data, showed better performance in differentiating between low and high nanomolar inhibitors. Therefore, it has been observed that all the hit rate profiles, for all four activity ranges and for all scoring schemes, for both programs, are significantly better than the random ones, confirming that the docking can be used successfully for enriching composite collections.

The study performed by Liu *et al*.[87] used PCA as an alternative method, to perform the combination of multiple punctuation functions, modifying punctuation, PCscore, re-scoring and re-ranking of compounds after VS. The main objective of this work was to identify possible inhibitors that are effective and selective of β-secretase (BACE-1), which may be more efficient than existing drugs for the treatment of Alzheimer's Disease (AD). The methodology of this work makes use of a training set that is formed by 50 inhibitors of BACE-1 and 9950 diverse inactive compounds. Three VS methods based on the individual score, conventional consensus score and PCA were used, which were evaluated by the number of hits in the first 1% of the classification. Then, the docking poses were generated by the Surflex program[87] and five scoring functions (Surflex_Score, D_Score, G_Score, ChemScore and PMF_Score) were used to extract the poses. For each pose group, 12

scoring functions (Surflex_Score, D_Score, G_Score, ChemScore, PMF_Score, LigScore1, LigScore2, PLP1, PLP2, Jain, Ludi_1, and Ludi_2) were used. 113,228 chemical compounds (Sigma-Aldrich® corporate chemical directory) were coupled by Surflex. When analyzing the training set, the PCA approach provided a superior rating when compared with the conventional consensus score and the single score. In the test set, the top 20 compounds that were chosen according to the conventional consensus score were submitted to experimental tests, but no inhibitor was found. Subsequently, they used the PCA scoring protocol to test 20 different compounds and it was through a fluorescence resonance energy transfer (FRET) assay that resulted in two drugs and BACE-1 inhibitors.

6. SCORING STRATEGIES CONSENSUS

6.1 Exponential Consensus Ranking (ECR)

A new consensus scorecard published recently[89] combines results from various docking programs using an exponential distribution for each individual classification. The theory consists in assigning an exponential punctuation p (r_i^j) for each molecule (*i*) for each scoring function (*j*) using the rank of the molecule (r_i^j) given by each individual anchor program,

$$p(r_i^j) = 1/\sigma \exp(-r_i^j/\sigma) \qquad (9)$$

where σ is the expected value of the exponential distribution. The final score of each molecule *i* is defined as the sum of the exponential punctuation of all punctuation functions *j*

$$P(i)=\sum_j p(r_i^j) = 1/\sigma \times \sum_j \exp \times (-r_i^j/\sigma) \qquad (10)$$

According to the authors, the new method exceeds the best traditional consensus strategies and is independent of scoring units, scales and compensations; preventing the combination of different program results (Fig. **3**).

6.2 REPPose

Ren *et al.* [21], also proposed a recently published strategy for selecting the best combinations of docking programs and improve hit rate using a method similar to 3DScore (VoteDock)[90]. For this, 3,535 protein-binding complexes served as input to 11 programs

namely, UCDOCK[29], UCSF DOCK, GalaxyDock, [91]–[93] LeDock,[94] MDock,[95] PLANTS, PSOVina,[96] QuickVina,[97] rDock,[98] Vina and Smina. [39]

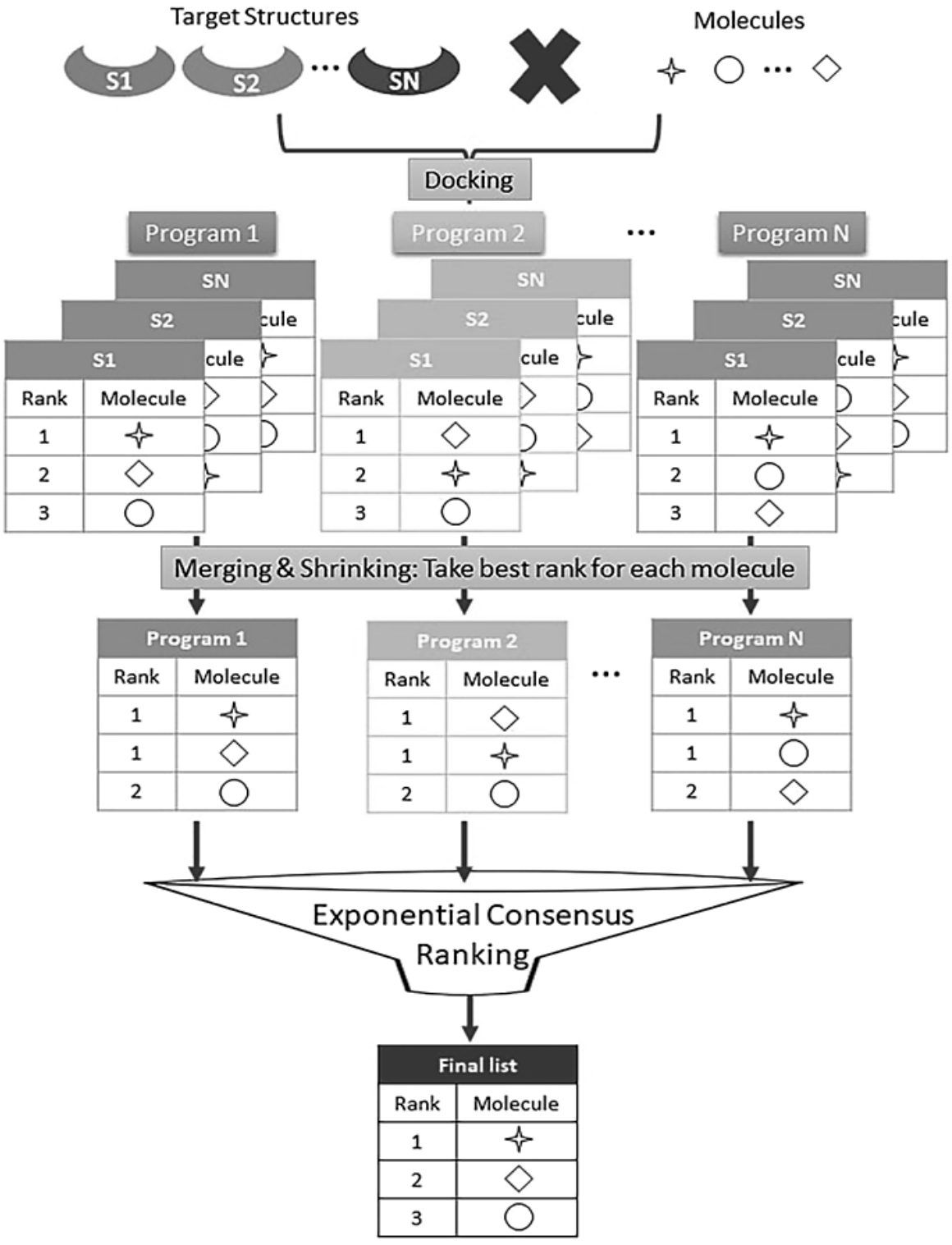


Fig. (3). Workflow using MS and ECR strategies. First, perform VS using different receiver structures and different docking programs/punctuation functions. Then apply the MS procedure, for each program, maintaining the best classification of each molecule. Finally, apply ECR as a method of consensus scoring (Source: Palacio-Rodríguez *et al*.[89]).

The method consists of representative pose determination (REPPose), i.e., after docking, the best scores of all programs are collected and used to determine REPPose. For each binder with *n* poses, the total RMSD (tRMSD) for each pose with the other poses is calculated, and the REPPose is defined as the pose with the minimum tRMSD. To determine the combination of the program the data of n higher punctuation poses generated from n programs, there are m non-repetitive program combinations (COMBs). As shown in Eq. 11, m is equal to all combinations of n poses without repetition. The term i starts from 3 because a REPPose can only be determined by at least 3 poses.

$$m = \sum_i (n/i),\ i = 3 \ldots \qquad (11)$$

Then, the COMBs are classified by a success rate of RMSD (≤2.0 Å) and the average of RMSD corresponding to "best" COMB (which have higher success rates or lower RMSD), are called COMBBestr and COMBBestm respectively. The authors concluded that there was a 4.9% increase in success rate compared with the best program.

6.3 Combined PLANTS-IFP

The method developed by Kooistra *et al*. [99] combines an anchor score function with molecular interaction fingerprint (IFP) to identify new ligands. In this study, a set of compounds were submitted to the docking using PLANTS selecting the ChemPLP scoring function. The IFP consists of evaluating the binding mode of a compound in a protein structure noting the absence or presence of different types of interactions (hydrophobic, aromatic, H-ionic, ionic) between each residue and the molecule based on a set of rules.[100] Thus, the marked reference IFPs are used to score the poses. For the validation of the method, 63 compounds were tested experimentally. The authors found that although the combined score approach was the most successful (between 53 and 73% accuracy), individual approaches also resulted in high hit rates.

6.4 SDC (Standard Deviation Consensus) and vSDC (variable Standard Deviation Consensus)

The standard deviation consensus (SDC) method consists of intersecting sets of molecules from various VS programs, based on the standard deviations of rank distributions, to rationalize the cutoff point. This is done by plotting the docking results for each program as ranking curves reporting the scores of the molecules versus their rank. The

histograms correspond to the results superimposed on the classification curves (Fig. **4**). The best scores that stood out in the standard deviation (SD) of the mean of the curves define the cutoff point. This method is dependent on the performance of the docking programs. Therefore, to calculate the cutoff point, use x versus SD, where x = number of best-performing programs.[99]

The consensus standard deviation method of the variable (vSDC) consists of assembling n molecules that are common to several otherwise equivalent programs to bring these molecules from their position quite far (1700 on average) to the first rows (less than 50). The cutoff point is a variable number of SDs gradually decreased until the defined number of common molecules (nCM = n test) is found [100].

6.5 X-CSCORE

The X-CSCORE is a consensus score function developed by Wang, Lai and Wang[47] that combines three empirical scoring functions (HPScore, HMScore and HSScore) to reduce the errors of an individual score. This punctuation function considers factors responsible for the interaction like van der Waals, hydrogen bonding, deformation penalty and hydrophobic effect. In addition, X-CSCORE was calibrated using a set of complex protein linkers with experimentally determined structures and binding affinities through multivariate regression analysis. This consensus score function considerably improves the coupling accuracy when compared to the conventional force field calculation used for molecular coupling.[47]

The X-CSCORE can be used with the AutoDock program as a tool to generate possible binder conformations and the conformations are re-evaluated by X-CSCORE.

6.6 Specific method of receiver adjustment/rescue

Gupta *et al.* [102] have created a rescue method in which different punctuation functions are combined with molecular descriptors that contribute to the prediction of binding affinity. The approach was initially applied as an average of the AutoDock Vina and GOLD scores (Chemscore, Asp and Goldscore). For the normalization of the data of the scores of the software, two formulas were used one obtain values between a 0 and 1:

(For positive scores)

$$\textit{Normalized score} = (x - \min)/(\max - \min), \quad (12)$$

(For negative scores)

$$\textit{Normalized score} = 1 - (x - \min)/(\max - \min), \quad (13)$$

where x = corresponding score, max = maximum score and min = minimum score of the data set.

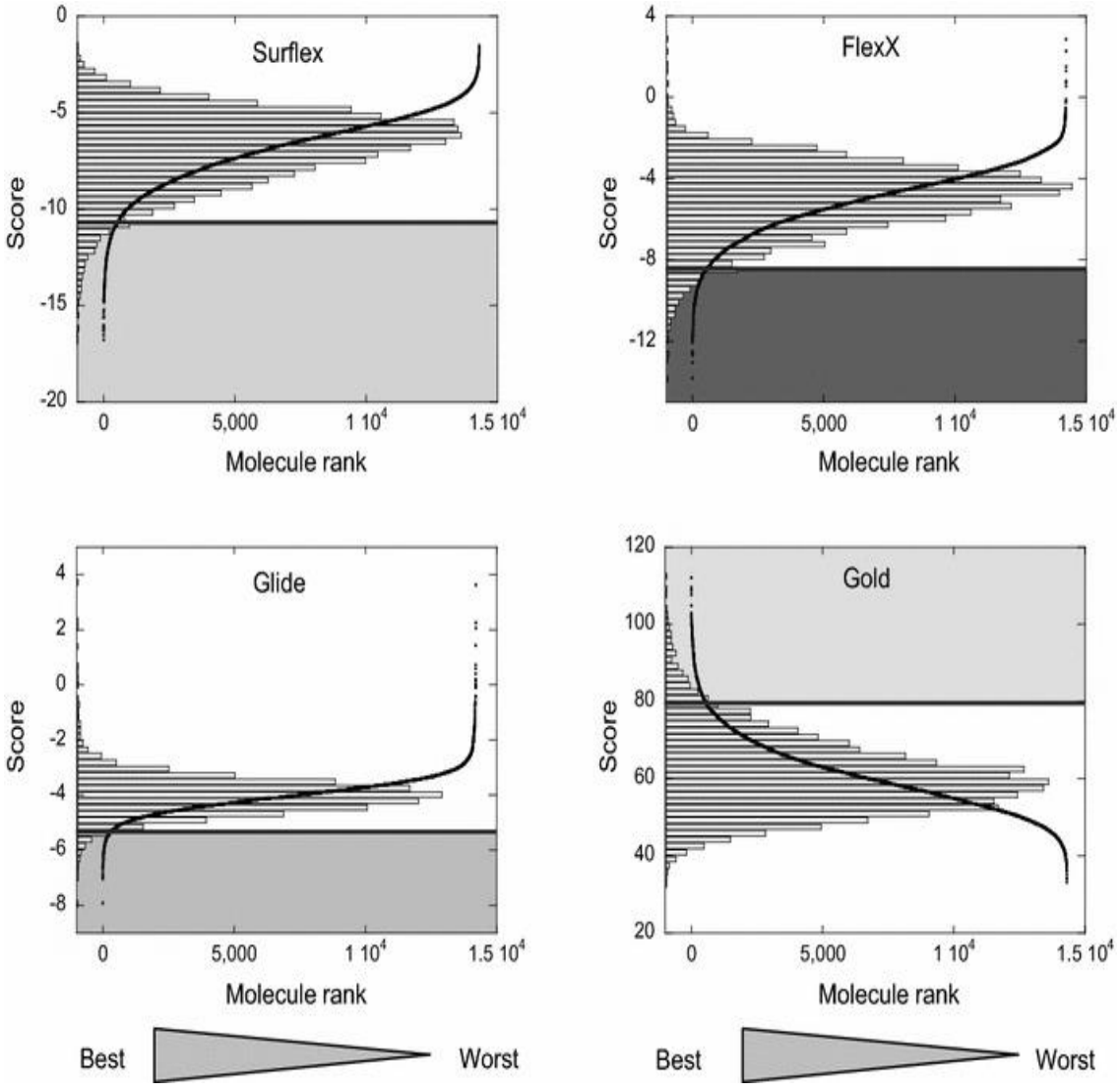


Fig. (4). The classification curves of the results of the VS obtained by the four programs. In these curves, the score obtained for each compound is positioned in relation to its classification given by the same program. The histogram of the punctuation distribution is superimposed on the classification curve. The horizontal line indicates the cutoff point of 2×SD that delimits between the molecules of greater prominence (in the colored boxes) (Source: Chaput *et al*.[101]).

Because contributes to protein-ligand and entropy interactions is mainly attributed to the hydrophobic and desolvation effects, the selected descriptors for the method were Log*P*, topological polar surface area (TPSA), inhibitor volume (Vol) and number of rotatable (Nrotb). It was also verified that of the descriptors considered, ΔGsolv and LogP showed higher correlations with the activity (pIC_{50}) of the compounds and rotatable bonds showed a negative correlation. The consensus scores also improved from 0.59 to 0.79. Thus, it is important to note that the method is dependent on the performance of the descriptors indicating the active compounds. Therefore, in this study, the formula that best represents the method is:

Re score = Docking score + LogP – Nrotb

6.7 Rank-by-number

Rank-by-number is obtained by using the sum of the punctuation values of each individual of the punctuation functions. The consensus score obtained by the sum total of the normalized value of the scoring functions is ranked according to the total obtained from the highest to the lowest score.[6]

6.8 Rank-by-rank

In the Rank-by-rank strategy, the normalized scores obtained are ranked based on their highest and lowest scores. A consensus ranking is then generated by adding the best ranks and eliminating the worst ranks for each scoring function. For example, if a molecule has 2, 4, 7, 8, 9, 14, and 17 classifications for seven different punctuation functions, the worst classifications (i.e., 9, 14 and 17) are dropped and finally the sum of the four (i.e., 2, 4, 7 and 8) is considered.[6]

6.9 Rank-by-vote

Rank-by-vote depends on the voting basis of the consensus score when the score value reaches a certain threshold criterion. This voting procedure is repeated for all the punctuation functions used to generate the consensus score, and the targets that have many votes are selected. Based on the threshold criteria, the voting classification approach is subdivided into (I) voting by number whose scores above the threshold criterion of 0.5 are

assigned one vote, (II) Vote-by-rank whose scores above row top limit criteria received a vote and (III) vote by whose scores are among the top 10% of the data set.[6]

6.10 VOTEDOCK

Plewczynski *et al*.[90] present a new docking methodology to select potent inhibitors using docking results performed by various programs. Thus, the main objective of this work was to present a new consensus approach to predict the structure of the protein-binding complex and its corresponding binding affinity. In this method, the results of seven docking programs (Surflex, LigandFit, Glide, GOLD, FlexX, eHiTS and AutoDock]) are used as inputs. To evaluate the method, the PDBbind database 200745-47 was composed of 3124 protein–ligand complexes with known 3D structure and binding affinity to the corresponding ligand. The quality of three consensus methods, MetaPose, MetaScore and VoteDock, was compared and compared with the results of the seven different docking programs. As a result, it was observed that the consensus algorithms are responsible for increasing the number of nesting pairs, with success of up to 70% for the VoteDock and 63% for the MetaPose, while the best fitting program of the evaluation reached less than 60% accuracy. It was also possible to observe the increase of the Pearson correlation of the predicted binding affinity compared with coefficient the experimental value to 0.5. The RMSD value of the predicted complex conformation versus its native conformation is reduced by a factor of 0.5 Å.

CONCLUSION

In conclusion, we note that the number of publications with consensus docking analyses has increased in recent years, although the first studies with this methodology were already reported in the literature for 20 years ago. Several consensus approaches have been developed and employed to increase the reliability of the results. Although docking is heavily used in virtual triages, their hit rates rarely exceed 60 percent. Thus, it is worrying when molecules from large virtual libraries are screened and many likely to be therapeutic potentials are discarded.

In several studies, it is possible to perceive the advantage of the consensus approach, which often exceeds 80% in the hit rate. However, the reliability of the docking consensus analysis depends not only on the number of samples tested, nor on the formulas used in the scoring methods but also on the good performance of the algorithm. For

example, if we used in a research five different programs docking to obtain a consensus score and if only one of the algorithm is responsible for the higher rate of accuracy compared with the others, when using the consensus score method, we risk decreasing rather than increasing the efficiency of the technique. In these cases, it is necessary to test and rank the algoritm that performs best before applying any consensus scoring strategy. For this, it is necessary to identify the type of algorithm and scoring function used by the software. This is because some algorithms and scoring functions disregard ligand size, number of rotary bonds, solvation energy, among others.

Punctuation strategies can be applied to poses or estimated binding energy values, even when results are obtained from punctuation functions based on force field calculations, empirical functions, and mean force knowledge. The important thing is to organize well the large amount of data generated and to know how to apply the consensus methods. Although the results of the analyzes are satisfactory using the consensus approach, it is still considered a robust method. Although the results of the analyses are satisfactory with the use of the consensus approach, it is still considered a robust method. Therefore, researchers working with multivariate statistical methodologies can program and contribute even more to the efficiency of the consensus docking, identifying not only active compounds of the inactive, as well as the properties and structures that favor the ligand potential, discriminating the best within a set of active molecules.

LIST OF ABBREVIATIONS

AD = Alzheimer's disease

AD4 = AutoDock 4

ADMET = Absorption, Distribution, Metabolism, Excretion and Toxicity

AMPK2 = α2 Subunit of monophosphate-activated protein kinase

BACE1 = Beta-secretase 1

CADD = Computer Aided Drug Design

CAIII = Carbonic Anhydrase III

CR = Rank of averages

DPP-4 = Dipeptidyl peptidase-4

DYNLL1 = Dynein light chain 1

FAAH = Fatty acid amide hydrolase

FBDD = Fragment-Based Lead Discovery

FDA = Food and Drug Administration

FLAP = Fingerprints for Ligands and Proteins

FRET = Fluorescence Resonance Energy Transfer

fXa = Coagulation enzyme factor Xa

FXR = Farnesoid X receptor

GC2 = Grand Challenge 2016

GWI = Gulf War Illness

H1N1 = Influenza A subtype

HA = Hemaglutinin

IC50 = Half maximal inhibitory concentration

IL-1Ra = Interleukin 1 receptor antagonist

IRHOM2 = Rhomboid protease 2

logP = Partition Coefficient

M2 = Ion channel

MADM = Mean absolute deviation from median

MbtI = Salicylate synthase of *Mycobacterium tuberculosis*

MIFs = Molecular interaction of fields

Mtb-DHFR = Dihydrofolate reductase of *Mycobacterium tuberculosis*

mTOR = Mammalian Target of Rapamycin

NA = Neuraminidase

NCI = National Cancer Institute

NP = Nucleoprotein

PA = Acidic polymerase protease

PCA = Principal Component Analysis

PDB = Protein Data Bank

PfFabD = Malonyl CoA-acyl carrier protein transacylase of *Plasmodium falciparum*

pIC50 = Logarithm of the maximal inhibitory concentration

PLS = Partial Least Squares

QSAR = Quantitative structure-activity relationships

RMSD = Root Mean Square Deviation

ROC = Receiver Operating Characteristic

SBP = Structure-based pharmacophore

SAR = Structure-activity relationship

SC = Score consensus

SDC = Standard Deviation Consensus

TgFabD = Malonyl CoA-acyl carrier protein transacylase of *Toxoplasma gondii*

TPSA = Topological polar surface area

VS = Virtual screening

vSDC = variable Standard Deviation Consensus

CONSENT FOR PUBLICATION

No applicable.

CONFLICT OF INTEREST

The authors declare no conflicts of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The present work was carried out with the support of the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

REFERENCES

(1) Gupta, M.; Sharma, R.; Kumar, A. Docking Techniques in Pharmacology: How Much Promising? *Comput. Biol. Chem.* **2018**, *76*, 210–217. https://doi.org/10.1016/j.compbiolchem.2018.06.005.

(2) Poli, G.; Martinelli, A.; Tuccinardi, T. Reliability Analysis and Optimization of the Consensus Docking Approach for the Development of Virtual Screening Studies. *J. Enzyme Inhib. Med. Chem.* **2016**, *31*, 167–173. https://doi.org/10.1080/14756366.2016.1193736.

(3) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **2012**, *7* (2), 146–157. https://doi.org/10.2174/157340911795677602.

(4) Onawole, A. T.; Kolapo, T. U.; Sulaiman, K. O.; Adegoke, R. O. Structure Based Virtual Screening of the Ebola Virus Trimeric Glycoprotein Using Consensus Scoring. *Comput. Biol. Chem.* **2018**, *72* (March 2016), 170–180. https://doi.org/10.1016/j.compbiolchem.2017.11.006.

(5) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42* (25), 5100–5109. https://doi.org/10.1021/jm990352k.

(6) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3–6), 1422–1426. https://doi.org/10.1021/ci010025x.

(7) Torres, P. H. M.; Sodero, A. C. R.; Jofily, P.; Silva-Jr, F. P. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20* (18), 4574. https://doi.org/10.3390/ijms20184574.

(8) Du, J.; Bleylevens, I. W. M.; Bitorina, A. V.; Wichapong, K.; Nicolaes, G. A. F. Optimization of Compound Ranking for Structure-Based Virtual Ligand Screening Using an Established FRED-Surflex Consensus Approach. *Chem. Biol. Drug Des.* **2014**, *83* (1), 37–51. https://doi.org/10.1111/cbdd.12202.

(9) Houston, D. R.; Walkinshaw, M. D. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. *J. Chem. Inf. Model.* **2013**, *53* (2), 384–390. https://doi.org/10.1021/ci300399w.

(10) Mavrogeni, M. E.; Pronios, F.; Zareifi, D.; Vasilakaki, S.; Lozach, O.; Alexopoulos, L.; Meijer, L.; Myrianthopoulos, V.; Mikros, E. A Facile Consensus Ranking Approach Enhances Virtual Screening Robustness and Identifies a Cell-Active DYRK1α Inhibitor. *Future Med. Chem.* **2018**, *10* (20), 2411–2430. https://doi.org/10.4155/fmc-2018-0198.

(11) Voet, A. R. D.; Kumar, A.; Berenger, F.; Zhang, K. Y. J. Combining in Silico and in Cerebro Approaches for Virtual Screening and Pose Prediction in SAMPL4. *J. Comput. Aided. Mol. Des.* **2014**, *28* (4), 363–373. https://doi.org/10.1007/s10822-013-9702-2.

(12) Bajusz, D.; Rácz, A.; Héberger, K. Comparison of Data Fusion Methods as Consensus Scores for Ensemble Docking. *Molecules* **2019**, *24* (15), 2690. https://doi.org/10.3390/molecules24152690.

(13) Bowen, L. R.; Li, D. J.; Nola, D. T.; Anderson, M. O.; Heying, M.; Groves, A. T.; Eagon, S. Identification of Potential Zika Virus NS2B-NS3 Protease Inhibitors via Docking, Molecular Dynamics and Consensus Scoring-Based Virtual Screening. *J. Mol. Model.* **2019**, *25* (7). https://doi.org/10.1007/s00894-019-4076-6.

(14) Salmas, R. E.; Seeman, P.; Aksoydan, B.; Erol, I.; Kantarcioglu, I.; Stein, M.; Yurtsever, M.; Durdagi, S. Analysis of the Glutamate Agonist LY404,039 Binding to Nonstatic Dopamine Receptor D2 Dimer Structures and Consensus Docking. *ACS Chem. Neurosci.* **2017**, *8* (6), 1404–1415. https://doi.org/10.1021/acschemneuro.7b00070.

(15) Wang, D.; Cui, C.; Ding, X.; Xiong, Z.; Zheng, M.; Luo, X.; Jiang, H.; Chen, K. Improving the Virtual Screening Ability of Target-Specific Scoring Functions Using Deep Learning Methods. *Front. Pharmacol.* **2019**, *10* (August), 1–11. https://doi.org/10.3389/fphar.2019.00924.

(16) Kalid, O.; Toledo, D.; Shechter, S.; Sherman, W.; Shacham, S. Consensus Induced Fit Docking (CIFD): Methodology , Validation , and Application to the Discovery of Novel Crm1 Inhibitors. *J. Comput. Aided. Mol. Des.* **2012**, 1217–1228. https://doi.org/10.1007/s10822-012-9611-9.

(17) Preto, J.; Gentile, F. Assessing and Improving the Performance of Consensus Docking Strategies Using the DockBox Package. *J. Comput. Aided. Mol. Des.* **2019**. https://doi.org/10.1007/s10822-019-00227-7.

(18) Perez-Castillo, Y.; Sotomayor-Burneo, S.; Jimenes-Vargas, K.; Gonzalez-Rodriguez, M.; Cruz-Monteagudo, M.; Armijos-Jaramillo, V.; Cordeiro, M. N. D. S.; Borges, F.; Sánchez-Rodríguez, A.; Tejera, E. CompScore: Boosting Structure-Based Virtual Screening Performance by Incorporating Docking Scoring Function Components into Consensus Scoring. *J. Chem. Inf. Model.* **2019**, *59* (9), 3655–3666.

https://doi.org/10.1021/acs.jcim.9b00343.

(19) Vilar, S.; Costanzi, S. Predicting the Biological Activities Through QSAR Analysis and Docking-Based Scoring. *Methods Mol. Biol.* **2012**, *914*, 271–284. https://doi.org/10.1007/978-1-62703-023-6_16,.

(20) Chermak, E.; De Donato, R.; Lensink, M. F.; Petta, A.; Serra, L.; Scarano, V.; Cavallo, L.; Oliva, R. Introducing a Clustering Step in a Consensus Approach for the Scoring of Protein-Protein Docking Models. *PLoS One* **2016**, *11* (11), 1–15. https://doi.org/10.1371/journal.pone.0166460.

(21) Ren, X.; Shi, Y. S.; Zhang, Y.; Liu, B.; Zhang, L. H.; Peng, Y. B.; Zeng, R. Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *J. Chem. Inf. Model.* **2018**, *58* (8), 1662–1668. https://doi.org/10.1021/acs.jcim.8b00329.

(22) Spinello, A.; Vecile, E.; Abbate, A.; Dobrina, A.; Magistrato, A. How Can Interleukin-1 Receptor Antagonist Modulate Distinct Cell Death Pathways? *J. Chem. Inf. Model.* **2019**, *59* (1), 351–359. https://doi.org/10.1021/acs.jcim.8b00565.

(23) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B. H.; Vreven, T.; Weng, Z. ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. *Bioinformatics* **2014**, *30* (12), 1771–1773. https://doi.org/10.1093/bioinformatics/btu097.

(24) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein-Protein Docking. *Nat. Protoc.* **2017**, *12* (2), 255–278. https://doi.org/10.1038/nprot.2016.169.

(25) Jiménez-García, B.; Pons, C.; Fernández-Recio, J. PyDockWEB: A Web Server for Rigid-Body Protein-Protein Docking Using Electrostatics and Desolvation Scoring. *Bioinformatics* **2013**, *29* (13), 1698–1699. https://doi.org/10.1093/bioinformatics/btt262.

(26) Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S. Y. HDOCK: A Web Server for Protein-Protein and Protein-DNA/RNA Docking Based on a Hybrid Strategy. *Nucleic Acids Res.* **2017**, *45* (W1), W365–W373. https://doi.org/10.1093/nar/gkx407.

(27) Lyskov, S.; Gray, J. J. The RosettaDock Server for Local Protein-Protein Docking. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), 233–238. https://doi.org/10.1093/nar/gkn216.

(28) Kausar, S.; Asif, M.; Bibi, N.; Rashid, S. Comparative Molecular Docking Analysis of Cytoplasmic Dynein Light Chain DYNLL1 with Pilin to Explore the Molecular Mechanism of Pathogenesis Caused by Pseudomonas Aeruginosa PAO. *PLoS One* **2013**, *8* (10), 1–14. https://doi.org/10.1371/journal.pone.0076730.

(29) Morris, G. M.;Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R.K.; Goodsell,

D.S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(30) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Res.* **2005**, *33* (SUPPL. 2), 363–367. https://doi.org/10.1093/nar/gki481.

(31) Mashiach, E.; Schneidman-Duhovny, D.; Andrusier, N.; Nussinov, R.; Wolfson, H. J. FireDock: A Web Server for Fast Interaction Refinement in Molecular Docking. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), 229–232. https://doi.org/10.1093/nar/gkn186.

(32) Ravikant, D. V. S.; Elber, R. PIE-Efficient Filters and Coarse Grained Potentials for Unbound Protein-Protein Docking. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (2), 400–419. https://doi.org/10.1002/prot.22550.

(33) Mashiach, E.; Nussinov, R.; Wolfson, H. J. FiberDock: Flexible Induced-Fit Backbone Refinement in Molecular Docking. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (6), 1503–1519. https://doi.org/10.1002/prot.22668.

(34) Jaundoo, R.; Bohmann, J.; Gutierrez, G. E.; Klimas, N.; Broderick, G.; Craddock, T. J. A. Using a Consensus Docking Approach to Predict Adverse Drug Reactions in Combination Drug Therapies for Gulf War Illness. *Int. J. Mol. Sci.* **2018**, *19* (11), 1–23. https://doi.org/10.3390/ijms19113355.

(35) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(36) Repasky, M. P. ; Shelley, M.; Friesner, R. A. Flexible Ligand Docking with Glide. *Curr. Protoc. Bioinforma.* **2007**, 8-12. https://doi.org/10.1002/0471250953.bi0812s18.

(37) Howell, D. C. *Median Absolute Deviation*; John Wiley & Sons: Hoboken, NJ, USA, 2014. https://doi.org/doi.org/10.1002/9781118445112.stat06232.

(38) Leherte, L.; Petit, A.; Jacquemin, D.; Vercauteren, D. P.; Laurent, A. D. Investigating Cyclic Peptides Inhibiting CD2–CD58 Interactions through Molecular Dynamics and Molecular Docking Methods. *J. Comput. Aided. Mol. Des.* **2018**, *32* (11), 1295–1313. https://doi.org/10.1007/s10822-018-0172-4.

(39) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53* (8), 1893–1904. https://doi.org/10.1021/ci300604z.

(40) Jones, Gareth; Willett, Peter; Glen, Robert C.; Leach, Andrew R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.*

1997, *267* (3), 727–748. https://doi.org/10.1006/jmbi.1996.0897.

(41) Olsson, M. A.; García-Sosa, A. T.; Ryde, U. Binding Affinities of the Farnesoid X Receptor in the D3R Grand Challenge 2 Estimated by Free-Energy Perturbation and Docking. *J. Comput. Aided. Mol. Des.* **2018**, *32* (1), 211–224. https://doi.org/10.1007/s10822-017-0056-z.

(42) FXR experimental data for the D3R Grand Challenge 2.

(43) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Comput. Chem.* **2005**, *26* (9), 915–931. https://doi.org/10.1002/jcc.20222.

(44) Poli, G.; Giuntini, N.; Martinelli, A.; Tuccinardi, T. Application of a FLAP-Consensus Docking Mixed Strategy for the Identification of New Fatty Acid Amide Hydrolase Inhibitors. *J. Chem. Inf. Model.* **2015**, *55* (3), 667–675. https://doi.org/10.1021/ci5006806.

(45) Cui, Y. H.; Chen, J.; Xu, T.; Tian, H. L. Structure-Based Grafting and Identification of Kinase-Inhibitors to Target MTOR Signaling Pathway as Potential Therapeutics for Glioblastoma. *Comput. Biol. Chem.* **2015**, *54*, 57–65. https://doi.org/10.1016/j.compbiolchem.2015.01.001.

(46) Eldridge, MD; Murray, CW; Auton, TR; Paolini, GV; Mee, R. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J Comput Aided Mol Des* **1997**, *11* (5), 425–445.

(47) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput. Aided. Mol. Des.* **2002**, *16* (1), 11–26. https://doi.org/10.1023/A:1016357811882.

(48) Velec, HF; Gohlke, H; Klebe, G. DrugScore(CSD)-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of near-Native Ligand Poses and Better Affinity Prediction. *J Med Chem* **2005**, *48* (20), 6296–6303.

(49) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A Knowledge-Based Energy Function for Protein-Ligand, Protein-Protein, and Protein-DNA Complexes. *J. Med. Chem.* **2005**, *48* (7), 2325–2335. https://doi.org/10.1021/jm049314d.

(50) Meng, E. C.;Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-based Energy Evaluation. *J. Comput. Chem.* **1992**, *13* (4), 505–524. https://doi.org/https://doi.org/10.1002/jcc.540130412.

(51) Goodsell, DS; Morris, GM; Olson, A. Automated Docking of Flexible Ligands:

Applications of AutoDock. *J Mol Recognit* **1996**, *9* (1), 1–5.

(52) Saikia, S.; Bordoloi, M.; Sarmah, R.; Kolita, B. Antiviral Compound Screening, Peptide Designing, and Protein Network Construction of Influenza a Virus (Strain a/Puerto Rico/8/1934 H1N1). *Drug Dev. Res.* **2019**, *80* (1), 106–124. https://doi.org/10.1002/ddr.21475.

(53) Thomsen, R.; Christensen, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. *J. Med. Chem.* **2006**, *49* (11), 3315–3321. https://doi.org/10.1021/jm051197e.

(54) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96. https://doi.org/10.1021/ci800298z.

(55) Sharma, K.; Tanwar, O.; Sharma, S.; Ali, S.; Alam, M. M.; Zaman, M. S.; Akhter, M. Structural Comparison of Mtb-DHFR and h-DHFR for Design, Synthesis and Evaluation of Selective Non-Pteridine Analogues as Antitubercular Agents. *Bioorg. Chem.* **2018**, *80* (April), 319–333. https://doi.org/10.1016/j.bioorg.2018.04.022.

(56) Wishart, D. S. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34* (90001), D668–D672. https://doi.org/10.1093/nar/gkj067.

(57) Xing, J.; Yang, L.; Li, H.; Li, Q.; Zhao, L.; Wang, X.; Zhang, Y.; Zhou, M.; Zhou, J.; Zhang, H. European Journal of Medicinal Chemistry Identi Fi Cation of Anthranilamide Derivatives as Potential Factor Xa Inhibitors : Drug Design , Synthesis and Biological Evaluation. *Eur. J. Med. Chem.* **2015**, *95*, 388–399. https://doi.org/10.1016/j.ejmech.2015.03.052.

(58) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites. *J. Mol. Graph. Model.* **2003**, *21* (4), 289–307. https://doi.org/10.1016/S1093-3263(02)00164-X.

(59) Park, H.; Eom, J. W.; Kim, Y. H. Consensus Scoring Approach to Identify the Inhibitors of AMP-Activated Protein Kinase A2 with Virtual Screening. *J. Chem. Inf. Model.* **2014**, *54* (7), 2139–2146. https://doi.org/10.1021/ci500214e.

(60) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. **1996**, 470–489.

(61) Xing, J.; Li, Q.; Zhang, S.; Liu, H.; Zhao, L.; Cheng, H.; Zhang, Y.; Zhou, J.; Zhang, H. Identification of Dipeptidyl Peptidase IV Inhibitors: Virtual Screening, Synthesis and Biological Evaluation. *Chem. Biol. Drug Des.* **2014**, *84* (3), 364–377. https://doi.org/10.1111/cbdd.12327.

(62) Lodhi, S. S.; Farmer, R.; Singh, A. K.; Jaiswal, Y. K.; Wadhwa, G. 3D Structure Generation, Virtual Screening and Docking of Human Ras-Associated Binding (Rab3A) Protein Involved in Tumourigenesis. *Mol. Biol. Rep.* **2014**, *41* (6), 3951–3959. https://doi.org/10.1007/s11033-014-3263-x.

(63) Newcombe, J.; Chatzidaki, A.; Sheppard, T. D.; Topf, M.; Millar, N. S. Diversity of Nicotinic Acetylcholine Receptor Positive Allosteric Modulators Revealed by Mutagenesis and a Revised Structural Model. *Mol. Pharmacol.* **2018**, *93* (2), 128–140. https://doi.org/10.1124/mol.117.110551.

(64) Aliebrahimi, S.; Montasser Kouhsari, S.; Ostad, S. N.; Arab, S. S.; Karami, L. Identification of Phytochemicals Targeting C-Met Kinase Domain Using Consensus Docking and Molecular Dynamics Simulation Studies. *Cell Biochem. Biophys.* **2018**, *76* (1–2), 135–145. https://doi.org/10.1007/s12013-017-0821-6.

(65) Onawole, A. T.; Sulaiman, K. O.; Adegoke, R. O.; Kolapo, T. U. Identification of Potential Inhibitors against the Zika Virus Using Consensus Scoring. *J. Mol. Graph. Model.* **2017**, *73*, 54–61. https://doi.org/10.1016/j.jmgm.2017.01.018.

(66) Hassaan, A.E; Sigler, S.C; Ibrahim, M.T.; Lee, K.J.; Cichon, L.K.; Gary, B.D; Joshua C. Canzoner, J.C.; Piazza, G. A; Ashraf, H. A. Mining ZINC Database to Discover Potential Phosphodiesterase 9 Inhibitors Using Structure-Based Drug Design Approach. *Med. Chem. (Los. Angeles).* **2016**, *12* (5), 472–477. https://doi.org/10.2174/1573406412666151204002836.

(67) Shah, J. J.; Khedkar, V.; Coutinho, E. C.; Mohanraj, K. Design, Synthesis and Evaluation of Benzotriazole Derivatives as Novel Antifungal Agents. *Bioorganic Med. Chem. Lett.* **2015**, *25* (17), 3730–3737. https://doi.org/10.1016/j.bmcl.2015.06.025.

(68) Kiss, R.; Sandor, M.; Szalai, F. A. Http://Mcule.Com: A Public Web Service for Drug Discovery. *J. Cheminform.* **2012**, *4* (S1), P17. https://doi.org/10.1186/1758-2946-4-s1-p17.

(69) Cerqueira, N. M. F. S. A.; Gesto, D.; Oliveira, E. F.; Santos-Martins, D.; Brás, N. F.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Receptor-Based Virtual Screening Protocol for Drug Discovery. *Arch. Biochem. Biophys.* **2015**, *582*, 56–67. https://doi.org/10.1016/j.abb.2015.05.011.

(70) Gangopadhyay, A.; Chakraborty, H. J.; Datta, A. Targeting the Dengue β-OG with Serotype-Specific Alkaloid Virtual Leads. *J. Mol. Graph. Model.* **2017**, *73*, 129–142. https://doi.org/10.1016/j.jmgm.2017.02.018.

(71) Mamidi, A. S.; Arora, P.; Surolia, A. Multivariate PLS Modeling of Apicomplexan FabD-Ligand Interaction Space for Mapping Target-Specific Chemical Space and Pharmacophore Fingerprints. *PLoS One* **2015**, *10* (11), 1–23. https://doi.org/10.1371/journal.pone.0141674.

(72) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132–1156. https://doi.org/10.1002/jcc.23905.

(73) Huang, S.; Song, C.; Wang, X.; Zhang, G.; Wang, Y.; Jiang, X.; Sun, Q.; Huang, L.; Xiang, R.; Hu, Y.; et al. Discovery of New SIRT2 Inhibitors by Utilizing a Consensus Docking/Scoring Strategy and Structure-Activity Relationship Analysis. *J. Chem. Inf. Model.* **2017**, *57* (4), 669–679. https://doi.org/10.1021/acs.jcim.6b00714.

(74) Sattarinezhad, E.; Bordbar, A. K.; Fani, N. Virtual Screening of Piperine Analogs as Survivin Inhibitors and Their Molecular Interaction Analysis by Using Consensus Docking, MD Simulation, MMPB/GBSA and Alanine Scanning Techniques. *J. Biomol. Struct. Dyn.* **2017**, *35* (8), 1824–1832. https://doi.org/10.1080/07391102.2016.1196152.

(75) Ju, Y.; Li, Z.; Deng, Y.; Tong, A.; Zhou, L.; Luo, Y. Identification of Novel BACE1 Inhibitors by Combination of Pharmacophore Modeling, Structure-Based Design and In Vitro Assay. *Curr. Comput. Aided-Drug Des.* **2016**, *12* (1), 73–82. https://doi.org/10.2174/1573409912666160222113103.

(76) Berry, M.; Fielding, B. C.; Gamieldien, J. Potential Broad Spectrum Inhibitors of the Coronavirus 3CLpro: A Virtual Screening and Structure-Based Drug Design Study. *Viruses* **2015**, *7* (12), 6642–6660. https://doi.org/10.3390/v7122963.

(77) Wang, Y.; Ge, H.; Li, Y.; Xie, Y.; He, Y.; Xu, M.; Gu, Q.; Xu, J. Predicting Dual-Targeting Anti-Influenza Agents Using Multi-Models. *Mol. Divers.* **2015**, *19* (1), 123–134. https://doi.org/10.1007/s11030-014-9552-4.

(78) Pini, E.; Poli, G.; Tuccinardi, T.; Chiarelli, L. R.; Mori, M.; Gelain, A.; Costantino, L.; Villa, S.; Meneghetti, F.; Barlocco, D. New Chromane-Based Derivatives as Inhibitors of Mycobacterium Tuberculosis Salicylate Synthase (Mbti): Preliminary Biological Evaluation and Molecular Modeling Studies. *Molecules* **2018**, *23* (7). https://doi.org/10.3390/molecules23071506.

(79) Dutta, D.; Das, R.; Mandal, C.; Mandal, C. Structure-Based Kinase Profiling to Understand the Polypharmacological Behavior of Therapeutic Molecules. *J. Chem. Inf. Model.* **2018**, *58* (1), 68–89. https://doi.org/10.1021/acs.jcim.7b00227.

(80) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2006**, *46* (1), 380–391. https://doi.org/10.1021/ci050283k.

(81) Raj, U.; Kumar, H.; Varadwaj, P. 3D Structure Generation, Molecular Dynamics and Docking Studies of IRHOM2 Protein Involved in Cancer & Rheumatoid Arthritis. *Curr. Comput. Aided-Drug Des.* **2016**, *11* (4), 325–335. https://doi.org/10.2174/1573409912666151124234008.

(82) Fourches, D.; Politi, R.; Tropsha, A. Target-Specific Native/Decoy Pose Classifier Improves the Accuracy of Ligand Ranking in the CSAR 2013 Benchmark. *J. Chem. Inf. Model.* **2015**, *55* (1), 63–71. https://doi.org/10.1021/ci500519w.

(83) Li, D. D.; Meng, X. F.; Wang, Q.; Yu, P.; Zhao, L. G.; Zhang, Z. P.; Wang, Z. Z.; Xiao, W. Consensus Scoring Model for the Molecular Docking Study of MTOR Kinase Inhibitor. *J. Mol. Graph. Model.* **2018**, *79*, 81–87. https://doi.org/10.1016/j.jmgm.2017.11.003.

(84) Alzweiri, M.; Al-Balas, Q.; Al-Hiari, Y. Chromatographic Evaluation and QSAR Optimization for Benzoic Acid Analogues against Carbonic Anhydrase III. *J. Enzyme Inhib. Med. Chem.* **2015**, *30* (3), 420–429. https://doi.org/10.3109/14756366.2014.940939.

(85) Ericksen, S. S.; Wu, H.; Zhang, H.; Michael, L. A.; Newton, M. A.; Hoffmann, F. M.; Wildman, S. A. Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2017**, *57* (7), 1579–1590. https://doi.org/10.1021/acs.jcim.7b00153.

(86) Cotesta, S.; Giordanetto, F.; Trosset, J. Y.; Crivori, P.; Kroemer, R. T.; Stouten, P. F. W.; Vulpetti, A. Virtual Screening to Enrich a Compound Collection with CDK2 Inhibitors Using Docking, Scoring, and Composite Scoring Odels. *Proteins Struct. Funct. Genet.* **2005**, *60* (4), 629–643. https://doi.org/10.1002/prot.20473.

(87) Liu, S.; Fu, R.; Zhou, L. H.; Chen, S. P. Application of Consensus Scoring and Principal Component Analysis for Virtual Screening against β-Secretase (BACE-1). *PLoS One* **2012**, *7* (6). https://doi.org/10.1371/journal.pone.0038086.

(88) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46* (4), 499–511. https://doi.org/10.1021/jm020406h.

(89) Palacio-Rodríguez, K.; Lans, I.; Cavasotto, C. N.; Cossio, P. Exponential Consensus Ranking Improves the Outcome in Docking and Receptor Ensemble Docking. *Sci. Rep.* **2019**, *9* (1), 1–14. https://doi.org/10.1038/s41598-019-41594-3.

(90) Plewczynski, D; Łaźniewski, M. von G. M. R. L. G. K. VoteDock: Consensus Docking Method for Prediction of Protein-Ligand Interactions. *J Comput Chem* **2011**, *32* (4), 568–581. https://doi.org/10.1002/jcc.21642.

(91) Shin, W. H.; Seok, C. GalaxyDock: Protein-Ligand Docking with Flexible Protein Side-Chains. *J. Chem. Inf. Model.* **2012**, *52* (12), 3225–3232. https://doi.org/10.1021/ci300342z.

(92) Shin, W. H.; Lee, G. R.; Seok, C. Evaluation of GalaxyDock Based on the Community Structure-Activity Resource 2013 and 2014 Benchmark Studies. *J. Chem. Inf. Model.* **2016**, *56* (6), 988–995. https://doi.org/10.1021/acs.jcim.5b00309.

(93) Baek, M.; Shin, W. H.; Chung, H. W.; Seok, C. GalaxyDock BP2 Score: A Hybrid Scoring Function for Accurate Protein–Ligand Docking. *J. Comput. Aided. Mol. Des.* **2017**, *31* (7), 653–666. https://doi.org/10.1007/s10822-017-0030-9.

(94) Zhang, N.; Zhao, H. Enriching Screening Libraries with Bioactive Fragment Space. *Bioorganic Med. Chem. Lett.* **2016**, *26* (15), 3594–3597. https://doi.org/10.1016/j.bmcl.2016.06.013.

(95) Huang, S. Y.; Zou, X. Construction and Test of Ligand Decoy Sets Using MDock: Community Structure-Activity Resource Benchmarks for Binding Mode Prediction. *J. Chem. Inf. Model.* **2011**, *51* (9), 2107–2114. https://doi.org/10.1021/ci200080g.

(96) Ng, M. C. K.; Fong, S.; Siu, S. W. I. PSOVina: The Hybrid Particle Swarm Optimization Algorithm for Protein-Ligand Docking. *J. Bioinform. Comput. Biol.* **2015**, *13* (3), 1–18. https://doi.org/10.1142/S0219720015410073.

(97) Alhossary, A.; Handoko, S. D.; Mu, Y.; Kwoh, C. K. Fast, Accurate, and Reliable Molecular Docking with QuickVina 2. *Bioinformatics* **2015**, *31* (13), 2214–2216. https://doi.org/10.1093/bioinformatics/btv082.

(98) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. RDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10* (4), 1–7. https://doi.org/10.1371/journal.pcbi.1003571.

(99) Kooistra, A. J.; Vischer, H. F.; McNaught-Flores, D.; Leurs, R.; De Esch, I. J. P.; De Graaf, C. Function-Specific Virtual Screening for GPCR Ligands Using a Combined Scoring Method. *Sci. Rep.* **2016**, *6* (May), 1–21. https://doi.org/10.1038/srep28288.

(100) Marcou, G.; Rognan, D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* **2007**, *47* (1), 195–207. https://doi.org/10.1021/ci600342e.

(101) Chaput, L.; Martinez-Sanz, J.; Quiniou, E.; Rigolet, P.; Saettel, N.; Mouawad, L. VSDC: A Method to Improve Early Recognition in Virtual Screening When Limited Experimental Resources Are Available. *J. Cheminform.* **2016**, *8* (1), 1–18. https://doi.org/10.1186/s13321-016-0112-z.

(102) Gupta, A.; Chaudhary, N.; Kakularam, K. R.; Pallu, R.; Polamarasetty, A. The Augmenting Effects of Desolvation and Conformational Energy Terms on the Predictions of Docking Programs against MPGES-1. *PLoS One* **2015**, *10* (8), 1–16. https://doi.org/10.1371/journal.pone.0134472

ANEXOS

ANEXOS I

Material suplementar do artigo: Virtual screening and the in vitro assessment of the antileishmanial activity of lignans

Table S1. Lignans with good ADMET profiles.

	Pharmacokinetics							
ID	**GI[1]**	**BBB[2]**	**P-gp[3]**	**CYP1A2**	**CYP2C19**	**CYP2C9**	**CYP2D6**	**CYP3A4**
01	High	No	Yes	No	No	No	No	No
08	High	No	Yes	No	No	No	No	No
20	High	No	Yes	No	No	No	Yes	No
30	High	No	Yes	No	No	No	Yes	No
31	High	No	Yes	No	No	No	Yes	No
32	High	No	Yes	No	No	No	Yes	No
44	High	No	Yes	No	No	No	Yes	No
57	High	No	Yes	No	No	No	Yes	No
58	High	No	Yes	No	No	No	Yes	No
59	High	No	Yes	No	No	No	Yes	No
60	High	No	Yes	No	No	No	Yes	No
61	High	No	No	No	No	No	Yes	No
76	High	No	No	No	No	No	Yes	No
77	High	No	No	No	No	No	Yes	No
83	High	No	Yes	No	No	No	Yes	No
84	High	No	Yes	No	No	No	Yes	No
86	High	No	Yes	No	No	No	No	No
87	High	No	Yes	No	No	No	Yes	No
88	High	No	Yes	No	No	No	No	No
90	High	No	Yes	No	No	No	No	No
91	High	No	Yes	No	No	No	Yes	No
121	High	No	Yes	No	No	No	No	No
124	High	No	Yes	No	No	Yes	No	No
131	High	No	No	No	No	No	Yes	Yes
132	High	No	Yes	No	No	No	No	No
139	High	No	Yes	No	No	No	No	No
140	High	No	Yes	No	No	No	No	No
151	High	No	Yes	No	No	No	Yes	No
156	High	No	Yes	No	No	No	Yes	No
157	Low	No	Yes	No	No	No	No	No
158	Low	No	Yes	No	No	No	No	No
159	Low	No	Yes	No	No	No	No	No
160	Low	No	Yes	No	No	No	No	No

[1]GI: Gastrointestinal absorption
[2]BBB: Blood–brain barrier
[3]P-gp: P-glycoprotein

Table S2. Predictive assessment of lignan toxicity for the evaluated parameters.

ID	Mutagenic	Tumorigenic	Reproductive Effective	Irritant
1	No	No	No	No
8	No	No	No	No
20	No	No	No	No
30	No	No	No	No
31	No	No	No	No
32	No	No	No	No
44	No	Low	No	No
57	No	No	No	No
58	No	No	No	No
59	No	No	No	No
60	No	No	No	No
61	No	No	No	No
76	No	No	No	No
77	No	No	No	No
83	No	No	No	No
84	No	No	No	No
86	No	No	No	No
87	No	No	No	No
88	No	No	No	No
90	No	No	No	No
91	No	No	No	No
121	No	No	No	No
124	No	No	No	No
131	No	No	No	No
132	Low	Low	No	No
139	No	No	No	No
140	No	No	No	No
151	No	No	No	No
156	No	No	No	No
157	No	No	No	No
158	No	No	No	No
159	No	No	No	No
160	No	No	No	No

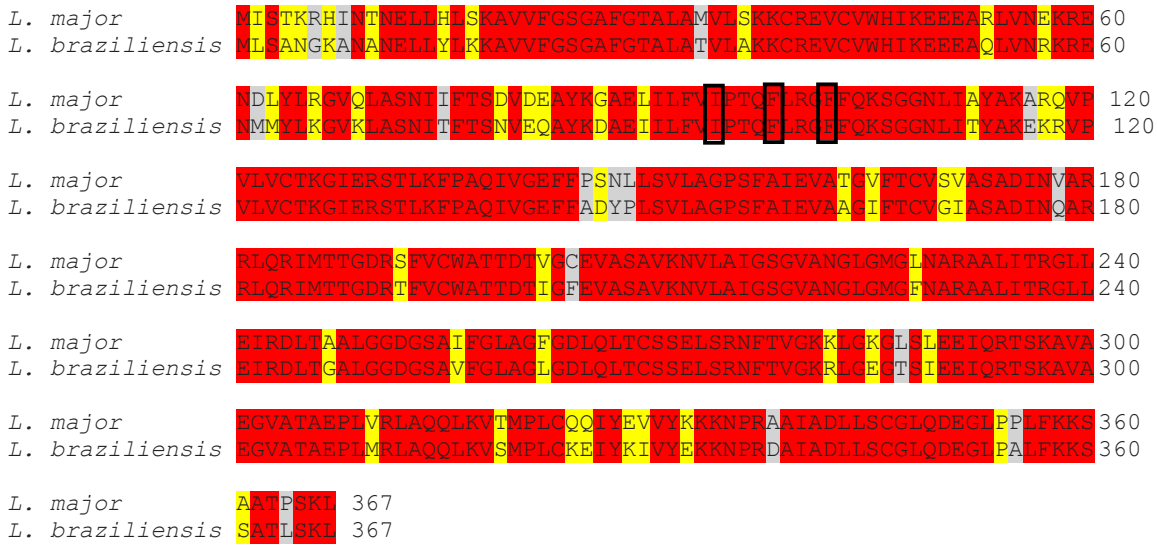


Figure S1. Alignment of the GPDH protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site.

Figure S2. Alignment of the DHODH protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site.

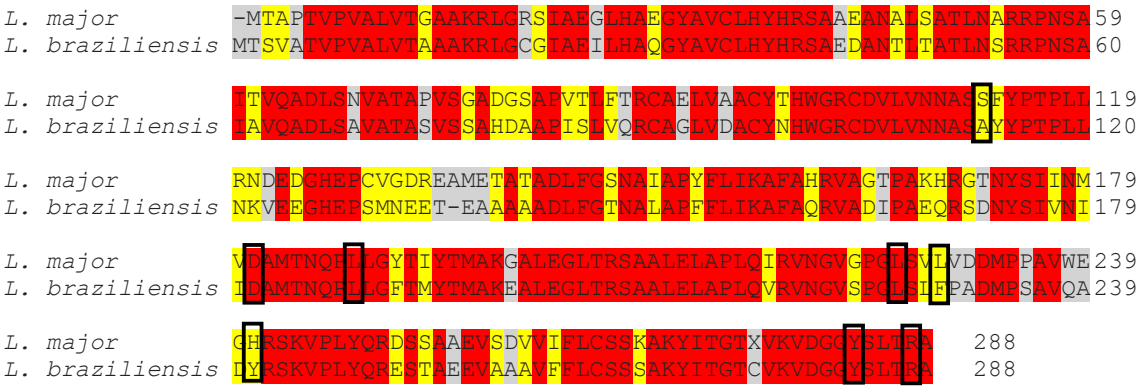


Figure S3. Alignment of the PTR1 protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site.

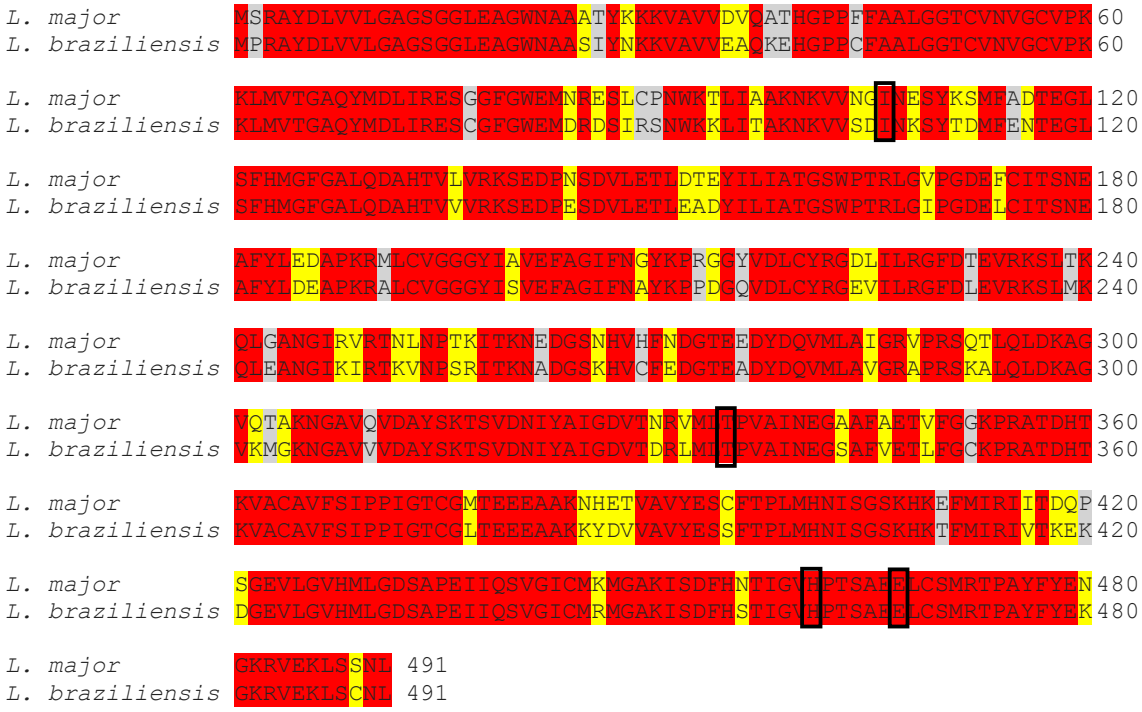


Figure S4. Alignment of the TR protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site.

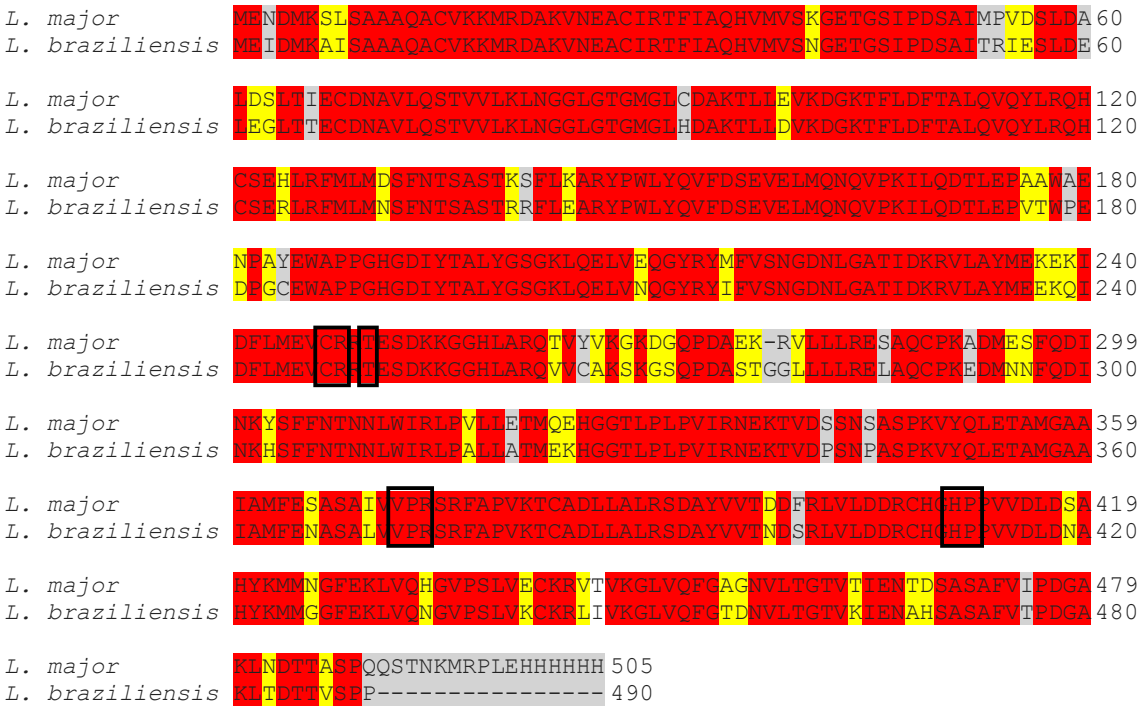


Figure S5. Alignment of the UGPase protein sequences. The gray regions show non-similar and non-identical amino acids. The red regions show identical amino acids. The yellow regions show similar amino acids. The black boxes indicate the active site.

Table S3. Average of all energy values (EM) obtained from the five scoring functions, for each lignan, and the probability value of potential consensus docking activity (P_{DC}), for each studied enzyme in *L. major*. Absent values indicate the molecules that were eliminated during this evaluation.

ID	GPDH		DHODH		PTR1		TR		UGPase	
	EM	P_{DC}	EM	P_{DC}	EM	P_{DC}	EM	P_{DC}	EM	P_{DC}
44	58.90	0.82	43.43	0.69	68.81	0.83	63.96	0.91	63.96	0.91
60	63.50	0.89	46.47	0.67	67.38	0.82	62.45	0.89	62.45	0.89
83	63.94	0.90	59.44	0.67	73.90	0.89	64.63	0.92	64.63	0.92
86	63.53	0.89	60.87	0.72	72.16	0.87	63.19	0.90	63.19	0.90
87	63.39	0.89	58.81	0.70	-	-	63.95	0.91	63.95	0.91
124	57.36	0.64	68.34	0.67	75.47	0.91	69.58	1	69.58	1
132	64.43	0.90	49.07	0.66	-	-	62.21	0.89	62.21	0.89
157	67.91	0.95	41.51	0.81	72.03	0.87	-	-	50.43	0.72
158	66.48	0.93	54.86	0.75	77.00	0.93	-	-	49.97	0.71
159	63.43	0.89	53.39	0.87	83.63	1	53.32	0.76	65.55	0.94
160	71.03	1	-	0.81	82.16	0.98	53.58	0.77	53.58	0.77
PDB inhibitor	37.72	0.53	35.75	0.58	63.94	0.77	37.38	0.53	37.38	0.53

Table S4. Average of all energy values (EM) obtained from the five scoring functions, for each lignan, and the probability value of potential consensus docking activity (P_{DC}), for each studied enzyme in *L. braziliensis*. Absent values indicate the molecules that were eliminated during this evaluation.

ID	GPDH		DHODH		PTR1		TR		UGPase	
	EM	**P_{DC}**	**EM**	**P_{DC}**	**EM**	**P_{DC}**	**EM**	**P_{DC}**	**EM**	**P_{DC}**
8	59.21	0.86	43.99	0.58	72.53	0.80	-	-	-	-
20	58.17	0.85	50.00	0.66	74.64	0.83	56.19	0.76	-	-
31	60.39	0.88	48.93	0.65	78.41	0.87	57.66	0.78	62	0.77
32	59.15	0.86	48.32	0.64	78.54	0.87	55.41	0.75	63	0.78
44	45.54	0.66	48.07	0.64	82.79	0.92	55.57	0.75	64	0.80
57	55.88	0.81	47.32	0.63	80.91	0.90	54.14	0.73	64	0.80
58	50.37	0.73	46.09	0.61	68.36	0.76	53.10	0.72	64	0.80
59	49.23	0.72	44.79	0.59	70.40	0.78	-	-	64	0.80
60	56.97	0.83	46.11	0.61	77.81	0.86	54.06	0.73	66	0.83
61	56.56	0.82	52.94	0.70	-	-	55.24	0.75	-	-
76	-	-	44.90	0.59	-	-	-	-	-	-
83	59.50	0.87	54.88	0.73	77.25	0.86	64.98	0.80	61	0.77
87	60.87	0.89	49.12	0.65	71.98	0.80	59.44	0.80	62	0.78
91	61.54	0.90	48.37	0.64	78.16	0.87	57.35	0.77	61	0.77
121	61.30	0.89	47.32	0.63	76.26	0.85	-	-	-	-
124	56.49	0.82	51.88	0.69	81.29	0.90	65.78	0.89	-	-
156	53.24	0.78	51.39	0.68	74.87	0.83	54.11	0.73	66	0.83
157	58.25	0.85	63.68	0.84	89.59	1	60.59	0.82	66	0.83
158	68.18	1	55.11	0.73	77.40	0.86	62.11	0.84	68	0.85
159	68.07	0.99	55.19	0.73	82.16	0.91	73.53	1	79	1
160	57.54	0.84	74.97	1	79.52	0.88	-	-	76	0.95
PDB inhibitor	36.26	0.53	32.40	0.43	67.31	0.75	53.15	0.72	61	0.76

ANEXOS II
Produção Científica

ARTIGOS COMPLETOS PUBLICADOS:

VIRTUAL SCREENING AND THE IN VITRO ASSESSMENT OF THE ANTILEISHMANIAL ACTIVITY OF LIGNANS

AUTORES: Mayara dos Santos Maia, Joanda Paolla Raimundo e Silva, Thaís Amanda de Lima Nunes, Julyanne Maria Saraiva de Sousa, Gabriela Cristina Soares Rodrigues, Alex France Messias Monteiro, Josean Fechine Tavares, Klinger Antonio da Franca Rodrigues, Francisco Jaime B. Mendonça-Junior, Luciana Scotti and Marcus Tullius Scotti.

REVISTA: ***Molecules*** 2020, 25(10), 2281; https://doi.org/10.3390/molecules25102281

Fator de impacto: 4.411

IDENTIFICATION OF NEW TARGETS AND THE VIRTUAL SCREENING OF LIGNANS AGAINST ALZHEIMER'S DISEASE

AUTORES: Mayara dos Santos Maia, Gabriela Cristina Soares Rodrigues, Natália Ferreira de Sousa, Marcus Tullius Scotti, Luciana Scotti, Francisco Jaime B. Mendonça-Junior.

REVISTA: ***Oxidative Medicine and Cellular Longevity,*** Volume 2020 |Article ID 3098673 | https://doi.org/10.1155/2020/3098673

Fator de impacto: 6.543

CONSENSUS ANALYSES IN MOLECULAR DOCKING STUDIES APPLIED TO MEDICINAL CHEMISTRY

AUTORES: Mayara dos Santos Maia, Gabriela Cristina Soares Rodrigues, Andreza Barbosa Silva Cavalcanti, Luciana Scotti, Marcus Tullius Scotti.

REVISTA: ***Mini-Reviews in Medicinal Chemistry,*** 2020;20(14):1322-1340, doi: 10.2174/1389557520666200204121129.

Fator de impacto: 3.862

LIGNANS AND NEOLIGNANS ANTI-TUBERCULOSIS IDENTIFIED BY QSAR AND MOLECULAR MODELING

AUTORES: Mayara S. Maia, Natália F. de Sousa, Gabriela C.S. Rodrigues, Alex F.M. Monteiro, Marcus T. Scotti, Luciana Scotti*

REVISTA: ***Combinatorial Chemistry & High Throughput Screening,*** 10.2174/1386207323666200226094940.

Fator de impacto: 1.339

LIGAND- AND STRUCTURE-BASED VIRTUAL SCREENING OF LAMIACEAE DITERPENES WITH POTENTIAL ACTIVITY AGAINST A NOVEL CORONAVIRUS (2019-NCOV)

AUTORES: Gabriela Cristina Soares Rodrigues, Mayara Dos Santos Maia, Renata Priscila Barros de Menezes, Andreza Barbosa Silva Cavalcanti, Natália Ferreira de Sousa, Érika Paiva de Moura, Alex France Messias Monteiro, Luciana Scotti, Marcus Tullius Scotti

REVISTA: ***Curr Top Med Chem*.** 2020;20(24):2126-2145. doi: 10.2174/1568026620666200716114546.

Fator de impacto: 3.295

QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP MODELING AND DOCKING OF MONOTERPENES WITH INSECTICIDAL ACTIVITY AGAINST *RETICULITERMES CHINENSIS* SNYDER AND *DROSOPHILA MELANOGASTER*

AUTORES: Gabriela Cristina Soares, Mayara dos Santos Maia, Eugene N. Muratov, Luciana Scotti, and Marcus Tullius Scotti*

REVISTA: ***J. Agric. Food Chem.*** 2020, 68, 16, 4687–4698, https://doi.org/10.1021/acs.jafc.0c00272.

Fator de impacto: 5.279

IN SILICO* STUDIES OF LAMIACEAE DITERPENES WITH BIOINSECTICIDE POTENTIAL AGAINST *APHIS GOSSYPII* AND *DROSOPHILA MELANOGASTER

AUTORES: Gabriela Cristina Soares Rodrigues, Mayara dos Santos Maia, Andreza Barbosa Cavalcanti, Marcus Tullius Scotti e Luciana Scotti.

REVISTA: ***Molecules*** 2021, *26*(3), 766; https://doi.org/10.3390/molecules26030766

Fator de impacto: 4.411

COMPUTER-ASSISTED DISCOVERY OF COMPOUNDS WITH INSECTICIDAL ACTIVITY AGAINST MUSCA DOMESTICA AND MYTHIMNA SEPARATA

AUTORES: Gabriela Cristina Soares Rodrigues, Renata Barros, Mayara dos Santos Maia, Andreza Barbosa Cavalcanti, Carlos L Cespedes-Acuña, Luciana Scotti e Marcus Tullius Scotti.

REVISTA: ***Food and Chemical Toxicology,*** DOI:10.1016/j.fct.2020.111899

Fator de impacto: 6.023

VIRTUAL SCREENING OF SECONDARYMETABOLITES OF THE FAMILY VELLOZIACEAE J. AGARDH WITH POTENTIAL ANTIMICROBIAL ACTIVITY

AUTORES: Anderson A. V. Pinheiro, Renata P. C. Barros, Edileuza B. de Assis, Mayara S. Maia, Diego I. A. F. de Araújo, Kaio A. Sales, Luciana Scotti, Josean F. Tavares, Marcus T. Scotti, Marcelo S. da Silva.

REVISTA: ***J. Braz. Chem. Soc.*** 31 (10), Oct 2020, https://doi.org/10.21577/0103-5053.20200112

Fator de impacto: 1.838

***IN SILICO* STUDY EXAMINING NEW PHENYLPROPANOIDS TARGETS WITH**

ANTIDEPRESSANT ACTIVITY

AUTORES: Poliane da Silva Calixto 1, Reinaldo Nóbrega de Almeida 2, Mirian G S Stiebbe Salvadori 3, Mayara Dos Santos Maia 4, José Maria Barbosa Filho 5, Marcus Tullius Scotti 4, Luciana Scotti 4.

REVISTA: ***Curr Drug Targets.*** 2021;22(5):539-554. doi: 10.2174/1389450121666200902171838.

Fator de impacto: 3.465

SELECTION OF ANTILEISHMANIAL SESQUITERPENELACTONES FROM SISTEMATX DATABASE USING A COMBINED LIGAND/STRUCTURE-BASED VIRTUAL SCREENING APPROACH

AUTORES: Chonny Herrera-Acevedo 1 2, Mayara Dos Santos Maia 1, Élida Batista Vieira Sousa Cavalcanti 1, Ericsson Coy-Barrera 2, Luciana Scotti 1, Marcus Tullius Scotti 3

REVISTA: Mol Divers. 2020 Sep 9. doi: 10.1007/s11030-020-10139-6.

Fator de impacto: 2.013

ANTINOCICEPTIVE ACTIVITY OF CHEMICAL COMPONENTS OF ESSENTIAL OILS THAT INVOLVES DOCKING STUDIES: A REVIEW

AUTORES: Davidson Barbosa Assis1, Humberto de Carvalho Aragão Neto1, Diogo Vilar da Fonsêca1, Humberto Hugo Nunes de Andrade1, Renan Marinho Braga1, Nader Badr2, Mayara dos Santos Maia3, Ricardo Dias Castro1, Luciana Scotti3, Marcus Tullius Scotti3* and Reinaldo Nóbrega de Almeida1

REVISTA: ***Front. Pharmacol.,*** 29 May 2020 | https://doi.org/10.3389/fphar.2020.00777

Fator de impacto: 7.768

DRUG DISCOVERY AND COMPUTATIONAL STRATEGIES IN THE MULTITARGETDRUGS ERA

AUTORES: Jéssika de Oliveira Viana, Mayara Barbalho Félix, Mayara dos Santos Maia, Vanessa de Lima Serafim, Luciana Scotti e Marcus Tullius Scotti.

REVISTA: ***Braz. J. Pharm. Sci.*** 54 (spe), 2018, https://doi.org/10.1590/s2175-97902018000001010

Fator de impacto: 1.838

A NEW LABDANE DITERPENE FROM THE AERIAL SEGMENTS OF LEPTOHYPTIS MACROSTACHYS (L'HÉRIT.) HARLEY & J.F.B. PASTORE

AUTORES: Andreza B.S.Cavalcanti, Pedro T.R.de Figueiredo, Carlos A.G.Veloso, Gabriela C.S.Rodrigues, Mayarados S. Maia, Alex France Messias Monteiro, Valnês S.Rodrigues Junior, Ana P.O.T.Castelo-Branco, Maria de F. Agra, Raimundo B.Filho, Marcelo S.da Silva, Josean F.Tavares, Vicente C.de O. Costa, Luciana Scotti, Marcus T.Scotti

REVISTA: ***Phytochemistry Letters,*** Volume 43, June 2021, Pages 117-122. https://doi.org/10.1016/j.phytol.2021.03.022

Fator de impacto: 1.679

ARTIGO ACEITO:

COMPUTER AIDED DRUG DESIGN METHODOLOGIES WITH NATURAL PRODUCTS IN THE DRUG RESEARCH AGAINST ALZHEIMER'S DISEASE

AUTORES: Natália Ferreira de Sousa a, Luciana Scotti a,b*, Érika Paiva de Moura a, Mayara dos Santos Maia a, Gabriela Cristina Soares Rodrigues a, Herbert Igor Rodrigues de Medeiros a, Simone Mendes Lopes and Marcus Tullius Scotti

REVISTA: ***Current Neuropharmacology***

Fator de impacto: 7.736