



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



CHONNY ALEXANDER HERRERA ACEVEDO

**Estudos quimiinformáticos da família Asteraceae na busca de estruturas com
potencial atividade contra doenças parasitárias negligenciadas.**

**Chemoinformatic studies of the Asteraceae family looking for structures with
potential activity against neglected parasitic diseases**

JOÃO PESSOA

2024

CHONNY ALEXANDER HERRERA ACEVEDO

Estudos quimioinformáticos da família Asteraceae na busca de estruturas com potencial atividade contra doenças parasitárias negligenciadas.

Chemoinformatic studies of the Asteraceae family looking for structures with potential activity against neglected parasitic diseases

*Tese de doutorado 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 obtenção do título de Doutor em Produtos Naturais e Sintéticos Bioativos, na área de concentração: **Farmacoquímica.***

Orientador: Prof. Dr. Marcus Tullius Scotti

Coorientador: Prof. Dr. Ericsson David Coy-Barrera

JOÃO PESSOA - PB

2024

Catálogo na publicação
Seção de Catalogação e Classificação

A174e Acevedo, Chonny Alexander Herrera.

Estudos quimioinformáticos da família Asteraceae na busca de estruturas com potencial atividade contra doenças parasitárias negligenciadas. / Chonny Alexander Herrera Acevedo. - João Pessoa, 2024.

254 f. : il.

Orientação: Marcus Tullius Scotti.

Coorientação: Ericsson David Coy-Barrera.

Tese (Doutorado) - UFPB/CCS.

1. Asteraceae - Leishmaniose. 2. Asteraceae - Estudos quimioinformáticos. 3. Leishmania. 4. Asteraceae - Metabolitos secundários. 5. Doenças tropicais negligenciadas. 6. Acoplamento molecular. I. Scotti, Marcus Tullius. II. Coy-Barrera, Ericsson David. III. Título.

UFPB/BC

CDU 582.998.1:616.993.161(043)

“La mente umana è inchinata naturalmente co' sensi a vedersi fuori nel corpo, e con molta difficoltà per mezzo della riflessione ad intendere se medesima.”

“A mente humana é naturalmente inclinada pelos sentidos a revelar-se fora no corpo, e com muita dificuldade, por meio da reflexão, a compreender-se a si mesma”

Giambattista Vico, *Scienza Nuova*

*A mi familia:
el amor y la gratitud por ustedes son infinitos.*

AGRADECIMENTOS

Agradeço ao CNPq pelo apoio financeiro mediante o processo 141341/2020-3.

À Universidade Federal da Paraíba e ao seu Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos: Coordenação, secretaria, professores, administrativos e estudantes.

Agradeço aos membros do laboratório de quimioinformática pelo trabalho em equipe, às colaborações e ao apoio constante durante esses anos.

À Renatinha, uma excelente pesquisadora e uma amiga maravilhosa. Minha melhor parceira de pesquisa, espero poder contar contigo sempre em qualquer parte do mundo.

Às amigas que fiz durante esses anos: Nikole Durand Trigueiro, admiro tua disciplina na pesquisa e amor pelos Produtos Naturais; Gleice Rayanne, não foi o jeito mais comum de começar a conversar, mas que bom que a vida permitiu te conhecer.

Ao Alixandre, pela amizade, por compartilhar o mesmo amor pelo futebol, mesmo não tendo as mesmas preferências. Obrigado por me ajudar nos momentos mais difíceis da pandemia; tua generosidade é uma das melhores coisas que conheci no Brasil. Sempre estarei em dívida contigo e com Cristiane.

À Luana Moraes, a quem considero uma amiga, agradeço muito pelo teu tempo, hospitalidade e disposição sempre, tua companhia constante como guia turística e principalmente por me levar ao circo.

À Professora Luciana Scotti, um exemplo de pesquisadora e de vida. A sabedoria da senhora sempre é uma luz na minha vida pessoal e profissional.

A todos os pesquisadores com quem fiz colaborações durante esses anos, agradeço a confiança no meu trabalho.

Ao Professor Marcus, meu mestre e amigo. Uma grande parte do que sou na minha vida é graças ao senhor. Seu apoio constante, pessoal e profissional, durante esses anos tem sido incrível, e só posso dizer que lhe admiro profundamente e que espero poder seguir trabalhando com o senhor e desfrutando de todos os campeonatos que estão por vir para os times que torcemos. Quando escrevo estas palavras, o Napoli ainda é o campeão da Itália e aqui está o verdadeiro epígrafe deste trabalho:

Sarò con te, Tu non devi mollare, Abbiamo un sogno nel cuore, Napoli torna campione.

AGRADECIMIENTOS

A mis padres por siempre enseñarme el valor que tienen las cosas, a partir de la dedicación y el trabajo, por su amor incondicional y ser mi ejemplo de vida, por todos sus sacrificios para poder llegar hasta el día de hoy.

A mis hermanos: Nury, mi compañera de vida; Nery: la alegría y el talento; Kenyi: el mejor Químico del mundo.

A todos mis tíos y primos porque son siempre una motivación y todo lo que hemos conseguido es también por el esfuerzo de ustedes.

A mis amigos, toda mi admiración y cariño por ustedes: Efraín, Felipe, Jorge, César Reyes, Tatiana, Joan, César Machuca, Mauricio, Karen, Francis, Sandra, Camilo Guerrero, Ángela Hernández, Adriana Barreto, María José Cortés siempre he estado acompañado gracias a ustedes, son alegría y fortaleza en este camino.

A Camilo Perdomo-Madrigal, mi coautor favorito y mi gran amigo, solo puedo darle las gracias por siempre, incluso en la distancia, estar conmigo.

A mi amiga querida: Ángela Naranjo, por caminar la mitad de la vida conmigo (hasta el momento) y confiar tanto en mí como persona y como profesional.

A Ángela María Moscote, una de las personas más incondicionales que he conocido y que más admiro en el mundo, eres luz y sabiduría, un ser increíble.

A la Dra. Claudia Vargas y el Dr. Leonardo Castellanos, gracias por su apoyo.

A la Universidad ECCI, en especial al Dr. Hernando Curtidor Castellanos por su apoyo y confiar siempre en mi trabajo profesional, así como a todos mis estudiantes que día a día me enseñan a ser un mejor químico.

A Ericsson Coy-Barrera, por su amistad de décadas, por todo el conocimiento que me ha transmitido, por su confianza en mi trabajo, por tanta paciencia y ser mi apoyo en Colombia, toda mi admiración y cariño, es un honor trabajar con usted y espero poder continuar aprendiendo de su disciplina y talento para la química.

Yenifer Sánchez Mesa, gracias por todo tu amor y compañía. Lo que pasamos durante estos años ha sido surreal, desde conocernos hasta empezar una familia, viviendo juntos en una pandemia. En lo efímero de la vida espero poder continuar a tu lado y buscar la felicidad juntos.

A Catalasa (EC 1.11.1.6), mi Paraibana favorita, gracias por elegirme; Luana Sofia, seja bem-vinda ao seu lar!

RESUMO

A leishmaniose, classificada como uma Doença Tropical Negligenciada (NTD), representa um desafio significativo para a saúde pública nas Américas. A forma clínica cutânea (CL) da doença tem um impacto considerável na América do Sul, com Brasil, Colômbia e Peru figurando entre os nove países que relatam 85% dos casos anuais em todo o mundo. O Brasil também enfrenta uma carga significativa de Leishmaniose Visceral (VL), a forma mais letal da doença. Apesar dos esforços de controle implementados nos últimos anos, a transmissão persiste em regiões pobres com condições precárias de higiene, facilitando o contato com os vetores da doença. Os tratamentos atuais, como compostos antimoniais, isetionato de pentamidina, miltefosina e anfotericina B liposomal, oferecem opções terapêuticas, mas não conseguem erradicar a infecção e apresentam diversos eventos adversos. Nesse contexto, surge a necessidade de desenvolver novas quimioterapias contra as diferentes formas clínicas da leishmaniose. Considerando a rica biodiversidade de Colômbia e Brasil, a busca por metabólitos secundários com propriedades leishmanicidas se torna relevante. A família botânica Asteraceae, com sua abundância de espécies (mais de 32.000) e uma ampla gama de metabólitos secundários, incluindo terpenoides e flavonoides, se apresenta como uma fonte promissória de moléculas bioativas. Atualmente, a seleção de moléculas com potencial atividade biológica é apoiada por ferramentas computacionais, demonstrando benefícios substanciais em termos de custo e tempo em comparação com abordagens clássicas, como a triagem em larga escala (High-throughput Screening). Esse enfoque permite uma exploração mais eficiente de compostos naturais para o desenvolvimento de novos tratamentos contra a leishmaniose.

Assim, este trabalho buscou realizar estudos quimioinformáticos da família Asteraceae para encontrar novas estruturas com potencial terapêutico contra as diversas formas clínicas da leishmaniose. O estudo começa com um capítulo dedicado aos paradigmas da descoberta de medicamentos com base nas estruturas do alvo terapêutico. Em seguida, no capítulo II, é realizada uma revisão de vários estudos que utilizaram ferramentas computacionais para examinar diversos compostos identificados na família Asteraceae na busca por possíveis candidatos a medicamentos contra *Leishmania*, destacando o uso de bancos de dados como uma ferramenta-chave na busca por potenciais moléculas leishmanicidas. Os capítulos subsequentes focam em estudos computacionais, utilizando diferentes abordagens de triagem virtual, utilizando diversas classes de metabólitos secundários presentes na Asteraceae para selecionar potenciais leishmanicidas. O capítulo III utiliza um banco de sesquiterpenlactonas presentes no SistemX para a seleção de potenciais leishmanicidas, empregando uma abordagem combinada de triagem virtual baseada na estrutura do ligante assim como do receptor, identificando estruturas com potencial inibitório para múltiplos alvos. O capítulo IV, dividido em duas partes, avalia a inibição da Pteridina redutase 1 (PTR1) e diidrofolato redutase – timidilato sintase (DHFR-TS), duas enzimas cruciais para o metabolismo desses parasitas tripanosomatídeos. Dois compostos com atividade inibidora dual para múltiplas espécies de *Leishmania* foram identificados partindo de um banco composto por 360 diterpenoides. Finalmente, no capítulo V, foram selecionadas lignanas do tipo butirolactona híbrida C6C3 como inibidores da *LmDHFR-TS* entre um banco de 314 derivados de ácido cinâmico, validando os modelos computacionais construídos mediante ensaios in-vitro com a enzima recombinante de DHFR-TS.

Palavras-chave: Asteraceae, triagem virtual, *Leishmania*, metabolitos secundários, aprendizado de máquina, doenças tropicais negligenciadas, acoplamento molecular.

ABSTRACT

Leishmaniasis, classified as a Neglected Tropical Disease (NTD), poses a significant challenge to public health in the Americas. The cutaneous form (CL) of the disease has a considerable impact in South America, with Brazil, Colombia, and Peru among the nine countries reporting 85% of annual cases worldwide. Brazil also faces a significant burden of Visceral Leishmaniasis, the most lethal form of the disease. Despite control efforts in recent years, transmission persists in impoverished regions with poor hygiene conditions, facilitating contact with disease vectors. Current treatments, such as antimonial compounds, pentamidine isethionate, miltefosine, and liposomal amphotericin B, offer therapeutic options but fail to eradicate the infection and present various adverse events. In this context, there is a need to develop new chemotherapies against the different clinical forms of leishmaniasis. Considering the rich biodiversity of Colombia and Brazil, the search for secondary metabolites with leishmanicidal properties becomes relevant. The botanical family Asteraceae, with its abundance of species (over 32,000) and a wide range of secondary metabolites, including terpenoids and flavonoids, emerges as a promising source of bioactive molecules. Currently, the selection of molecules with potential biological activity is supported by computational tools, demonstrating substantial benefits in terms of cost and time compared to classical approaches like High-throughput Screening. This approach allows for a more efficient exploration of natural compounds for the development of new treatments against leishmaniasis. This doctoral work aimed to conduct chemoinformatics studies of the Asteraceae family to find new structures with therapeutic potential against the various clinical forms of leishmaniasis. The study begins with a chapter dedicated to the paradigms of drug discovery based on the structures of the therapeutic target. Next, in Chapter II, a review of various studies that used computational tools to examine compounds identified in the Asteraceae family in the search for potential drug candidates against *Leishmania* is conducted, highlighting the use of databases as a key tool in the search for potential leishmanicidal molecules. Subsequent chapters focus on computational studies, using different virtual screening approaches and various classes of secondary metabolites present in Asteraceae to select potential leishmanicidal compounds. Chapter III uses a library of sesquiterpene lactones from Sistemax to select potential leishmanicidal compounds, employing a combined virtual screening approach based on ligand and receptor structure, identifying structures with inhibitory potential for multiple targets. Chapter IV, divided into two parts, evaluates the inhibition of Pteridine reductase 1 (PTR1) and dihydrofolate reductase-thymidylate synthase (DHFR-TS), two crucial enzymes for the metabolism of these trypanosomatid parasites. Two compounds with dual inhibitory activity for multiple *Leishmania* species were identified from a library of 360 diterpenoids. Finally, in Chapter V, hybrid C6C3 butyrolactone lignans were selected as inhibitors of *Lm*DHFR-TS from a library of 314 cinnamic acid derivatives, validating the computational models through *in vitro* assays with the recombinant DHFR-TS enzyme.

Keywords: Asteraceae, virtual screening, *Leishmania*, secondary metabolites, machine learning, neglected tropical diseases, molecular docking.

LISTA DE FIGURAS

Capítulo I

- Figure 1:** Examples of drugs commercially available that were developed from target-based drug design. 34
- Figure 2:** General workflow of the structure-based drug design (SBDD) 35

Capítulo II

- Figure 1:** Asteraceae is highly distributed worldwide. More than 23,000 species and 22 million georeferenced reports are registered in the Global Biodiversity Information Facility (GBIF). 70
- Figure 2:** Asteraceae secondary metabolites were identified as *hits* based on *in silico* methodologies. 75

Capítulo III

- Figure 1:** The VS methodology used in this study. Solid blue lines represent the two sets of compounds used to generate and validate the RF model for *L. donovani* amastigotes and promastigotes. 109
- Figure 2:** ROC plot, sensitivity versus (1 - specificity), generated by the selected RF model for cross-validation and 5-fold external validation of the promastigote and amastigote models. 112
- Figure 3:** Potentially active sesquiterpene lactones (five best-ranked), identified using ligand-based VS, for the two-parasitic forms of *L. donovani*. 113
- Figure 4:** Best-ranked antileishmanial SLs from the structure-based virtual screening. 117
- Figure 5:** Hydrogen-bonding interactions (blue dotted lines) and steric interactions (green dotted lines) between **a** pyridoxal-5-phosphate (PDB ID: PLP) and **b** disecoeudesmanolide 3A (structure **14**) at the active site of *L. donovani* ODC protein. 117
- c** An inhibitor reported in the PDB for MPK3 (PDB ID:046) and **d** disecoeudesmanolide 3A (structure **14**), in the active site of MPK3 protein.
- Figure 6:** Docking formations between **a** 5-sarracinate-eucannabinolide (structure **11**), **b** pannonicoside (structure **12**), and **c** 2-oxopentadecyl-CoA (PDB ID: NHW) at the active site of *L. donovani* NMT. 118
- Figure 7:** Representation of the best-ranked structures for each parasitic form, 120

identified using a combined ligand-based and structure-based virtual screening approach.

Figure 8: Root-mean-square differentiation (RMSD), **b** root-mean square fluctuation (RMSF), and **c** radius of gyration (RoG) values within the *LdPTR1* binding site, obtained after molecular dynamics simulations using five of the best-ranked SLs, identified using various VS approaches. 122

Capítulo IV

Parte A

Figure 1: Potentially active kauranes, identified using RF model (ligand-based VS), for *L. major*. 141

Figure 2: **a)** Docking conformations of structure **135** (green), **302** (red) and MTX (yellow) in the active site of *LmPTR1*; 2D-residual interaction diagrams of **(b)** methotrexate (MTX), **(c)** structure **135**, and **(d)** structure **3**. 144

Figure 3: Synthetic route to produce monoterpene/diterpene ester adduct **135**. 146

Figure 4: Synthetic route to produce phenylpropanoid/diterpene ester adducts **301**, **302**, **301a**, and **302a**. 147

Figure 5: Docking formations between a structure **302** in the active site of **(a)** *LpPTR1*, **(b)** *LbPTR1*, and **(c)** *LaPTR1*. 152

Figure 6: **(a)** Root-mean-square deviation (RMSD), **(b)** root-mean-square-fluctuation (RMSF), and **(c)** radius of gyration (RoG) values within the *LbPTR1* binding site, obtained after molecular dynamics simulations using three of the best-ranked structures in molecular docking calculations. 153

Parte B

Figure 1: **(a)** Structures of selected kaurane-type diterpenes (**135**, **301**, **302**) and their derivatives (**301a** and **302a**). **(b)** Synthesis of compound **4** from structure **148**. 176

Figure 2: Two-dimensional residual interaction diagrams in the active site of *L. major* DHFR-TS for: **(a)** structure **4**, **(b)** structure **135**, **(c)** structure **301**, **(d)** structure **302**, **(e)** structure **301a**, **(f)** structure **302a** and **(g)** methotrexate (MTX). 180

Figure 3: Best pose of structure **302** (orange) and MTX (red) in the active site of **(a–d)** *L. amazonensis* **(b–e)** *L. braziliensis* and **(c–f)** *L. panamensis* DHFR-TS (green). 182

Figure 4: **(a,b)** Root-mean-square deviation (RMSD), **(c,d)** root-mean-square fluctuation (RMSF), and **(e,f)** radius of gyration (RoG) values within the *L. major* DHFR-TS and *L. braziliensis* DHFR-TS binding site, obtained after molecular 184

dynamics simulations.

Capítulo V

Figure 1: ROC curves comparison for the RF model using AlvaDesc and VolSurf descriptors for **(a)** test sets and **(b)** cross-validation. Performance evaluation of RF using **(c)** AlvaDesc and **(d)**VolSurf descriptors. **(e)** Precision-recall (PR) curves for cross-validation. **(f)** Scatter plots depicting the results of the PCA analysis conducted on the training and test datasets. 205

Figure 2: Chemical Structures of the five top-ranked cinnamic acid derivatives using a ligand-based virtual screening (LB) with AlvaDesc and VolSurf+ descriptors. 208

Figure 3: **(b)** Chemical structures of reference ligands: Methotrexate (MTX) and (DQ1). Redocking results of **(b)** MTX and **(c)** DQ1 in the active site of *Lm*DHFR-TS. 210

Figure 4: Chemical Structure of six of the best-ranked cinnamic acid derivatives that appear as active in the structure-based virtual screening with their respective probability to be active. 211

Figure 5: Residual interaction diagrams of **(a)** compound **241**, **(b)** compound **164**, **(c)** compound **21**, **(d)** DQ1, and **(e)** methotrexate. 213

Figure 6: Topological Polar Surface Area (TPSA) map of **(a)** DQ1, **(b)** methotrexate, **(c)** compound **241**, **(d)** compound **164**, and **(e)** compound **21**. 214

Figure 7: Cinnamic acid derivatives as potential inhibitors of *Lm*DHFR-TS were identified using an approach that combines ligand-based and structure-based virtual screening (VS). 217

Figure 8. **(a)** Root-mean-square deviation (RMSD), **(b)** root-mean-square fluctuation (RMSF), and **(c)** radius of gyration (RoG) values within the *Lm*DHFR-TS binding site, obtained after molecular dynamics simulations. 219

LISTA DE TABELAS

Capítulo I

Table 1: Examples of binding site identification software classified according to the type of algorithm used for the prediction	39
Table 2: Summary of the most common sampling algorithms for molecular docking.	42
Table 3: Examples of the scoring functions in molecular docking calculations.	45
Table 4: Main classification of the most common MD algorithms.	48
Table 5: Examples of de novo software classified according to the type of methodology used to construct and score molecules.	52

Capítulo II

Table 1: Enzymes from different <i>Leishmania</i> species were studied by molecular docking, using secondary metabolites from Asteraceae.	80
--	----

Capítulo III

Table 1: Summary of cross-validation and 5-fold external validation results, which were obtained using the RF algorithm on the total set of 3,159 compounds for <i>L. donovani</i> amastigotes, and 1,569 compounds for <i>L. donovani</i> promastigotes	110
Table 2: The docking energy (kJ/mol) for the best-ranked SLs, according to the structure-based approach, for each of the four <i>L. donovani</i> proteins studied.	116
Table 3: Summary of the best-ranked structures, identified using a combined ligand and structure-based virtual screening approach.	121
Table 4: Binding free energies (kJ/mol) from the MM/PBSA calculations for five of the best-ranked structures identified for <i>LdPTR1</i>	124

Capítulo IV

Parte A

Table 1: Docking energies of the best-ranked structures from the structure-based VS for <i>L. major</i> PTR1.	143
Table 2: Kauranes classified as active using an approach combining ligand-based and structure-based VS.	145

Table 3: Results for VS-selected diterpenes as inhibitors of <i>Lm</i> PTR1.	148
Table 4: Kauranes classified as active using an approach combining ligand-based and structure-based VS.	150
Table 5: VINA scores and interactions of structure 302 , PMA and DHB with amino acid residues of <i>Lp</i> PTR1, <i>Lb</i> PTR1, and <i>La</i> PTR1.	151
Table 6: Binding free energies (kJ/mol) from the MM/PBSA calculations for three of the best-ranked structures identified for <i>Lb</i> PTR1; DHB and PMA were used as reference ligands.	154

Parte B

Table 1: Results of enzymatic activity against <i>L. major</i> dihydrofolate reductase (<i>Lm</i> DHFR) for selected kaurane-type diterpenes.	176
Table 2: MolDock scores of six kaurane-type diterpenes, MTX, and a 2,4-diamine derivative (a dual inhibitor of PTR1/DHFR-TS) against <i>L. major</i> DHFR-TS.	178
Table 3: The VINA score values for six tested structures and MTX (methotrexate) for dihydrofolate reductase-thymidylate synthase (DHFR-TS) of <i>Leishmania braziliensis</i> , <i>Leishmania panamensis</i> , and <i>Leishmania amazonensis</i> .	181
Table 4: Binding free energies (kJ/mol) from the MM/PBSA calculations for structure 302 and its derivative 302a for <i>L. major</i> DHFR-TS and <i>L. braziliensis</i> DHFR-TS; In both proteins MTX was used as reference ligand.	186

Capítulo V

Table 1: Chemical Structure of six of the best-ranked cinnamic acid derivatives that appear as active in the structure-based virtual screening with their respective probability to be active.	212
Table 2: Cinnamic acid derivatives are classified as active by combining ligand-based and structure-based VS.	216
Table 3: Binding free energies (kJ/mol) from the MM/PBSA calculations for compounds 237 , 306 , and 308 in the active site of <i>Lm</i> DHFR-TS; MTX was used as the reference ligand.	221
Table 4: Results of enzymatic activity against <i>Lm</i> DHFR-TS and <i>Hs</i> DHFR for selected cinnamic acid derivatives.	222

LISTA DE ABREVIATURAS, SIGLAS E FÓRMULAS

2D	Two-Dimensional
3D	Three-Dimensional
<i>A. annua</i>	<i>Artemisia annua</i>
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
APD	Applicability Domain
AUC	Area Under the ROC Curve
BLAST	Basic Local Alignment Search Tool
bpol	Polarizability
BSA	Bovine Serum Albumin
CA	Combined Probability Values
CADD	Computer-Aided Drug Design
CAST	Computed Atlas of Surface Topography
cathepsin B	Cathepsin B
ChEMBL	Chemical Biology Database
CL	Cutaneous leishmaniasis
Clarivate	A global information services company
COCONUT	Comprehensive Natural Products Database
ConSurf	Conserved Surface Residue
Cryo-EM	Cryogenic Electron Microscopy
CS	Cysteine Synthase
CSV	Comma-Separated Value
CV	Cross-validation
CYP	Cytochrome
CYP51	Sterol 14 α -demethylase
DBs	Databases
DEPTH	Database of Experimental Protein Structures in Trimeric Transmembrane
DHB	Dihydrobiopterin
DHF	7,8-Dihydrofolate
DHFR-TS	Dihydrofolate reductase - Thymidilate synthase
DHODH	Dihydroorotate Dehydrogenase

DL	Deep Learning
DNA	Deoxyribonucleic Acid
DNDD	De Novo Drug Design
DNP	Dictionary of Natural Products
DoGSite Scorer	A server for predicting ligand-binding sites
DRL	Deep Reinforcement Learning
EDTA	Ethylenediaminetetraacetic acid
E_i	Docking energy.
E_{ligand}	Ligand energy
E_{min}	Lowest energy value
ExPASy	Expert Protein Analysis System
FAK	Focal adhesion kinase
FASTA	Fast-All
FDA	United States Food and Drug Administration
FF	Force Field
FN	False-negative rate
FP	False-positive rate
G proteins	Guanine Nucleotide-Binding Proteins
GA	Genetic Algorithm
GBIF	Global Biodiversity
GP63	Leishmanolysin
GPS	Global Positioning System
GPU	Graphics Processing Unit
GROMACS	GRoningen MACHine for Chemical Simulations
H-bond	Hydrogen bond
HQSAR	Hologram QSAR
Hs	<i>Homo sapiens</i>
HsDHODH	Human DHODH
HTS	High-throughput screening
IC₅₀	Half-maximal Inhibitory Concentration
K_i	Inhibitory constant
KNAPSAcK	Kyoto Encyclopedia of Genes and Genomes (KEGG) Nucleic Acid Sequence Database

KNIME	Konstanz Information Miner
<i>L. amazonensis</i>	<i>Leishmania amazonensis</i>
<i>L. braziliensis</i>	<i>Leishmania braziliensis</i>
<i>L. donovani</i>	<i>Leishmania donovani</i>
<i>L. panamensis</i>	<i>Leishmania panamensis</i>
<i>L.major</i>	<i>Leishmania major</i>
<i>La</i>	<i>Leishmania amazonensis</i>
L_B	Ligand-based probability
<i>Lb</i>	<i>Leishmania braziliensis</i>
LBSS	Ligand-Based Virtual Screening
<i>Ld</i>	<i>Leishmania donovani</i>
LIGSITE	Ligand Binding Site Prediction
LISE	Ligand-Induced Structural Effects
<i>Lm</i>	<i>Leishmania major</i>
LmPTR1	<i>Leishmania major</i> Pteridine Reductase 1
LogP	Partition coefficient
<i>Lp</i>	<i>Leishmania panamensis</i>
MAP	Mitogen-Activated Protein
MCL	Mucocutaneous Leishmaniasis
MCC	Matthew's correlation coefficient
MCSS	Multiple Copy Simultaneous Search
MD	Molecular dynamics
Metapocket	A meta-approach for ligand-binding pocket prediction
MIF	Molecular Interaction Fields
MLP	Molecular Lipophilicity Potential
MLR	Multivariate Linear Regression
MM	Molecular Mechanics
MM/PBSA	Molecular Mechanics Poisson–Boltzmann Surface Area
MODELLER	Molecular Modeling Environment
MOE	Molecular Operating Environment
MPK3	Mitogen-Activated Protein Kinase 3
MS	Mass Spectrometry
MSM	Markov State Models

MTX	Methotrexate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NAPRALERT	Natural Products Alert
NMR	Nuclear Magnetic Resonance
NMT	N-myristoyltransferase
ns	nanoseconds
NSiteMatch	Nucleic Acid-Binding Site Match
NTD	Neglected Tropical Diseases
NUBBEdb	Natural Products in Brazil Database
ODC	Ornithine Decarboxylase
OPLS-DA	Orthogonal Projections to Latent Structures Discriminant Analysis
PAINS	Pan Assay Interference Compounds
PCA	Principal Component Analysis
PDB	Protein Data Bank
pIC₅₀	Negative logarithm (base 10) of the IC ₅₀ values
pKa	Acid dissociation constant values
PLS	Partial Least-Squares
PME	Particle-mesh Ewald
POVME	Pocket Volume Measurer
PRC	Precision-recall curve
PROCHECK	Program to Check the Stereochemical Quality of a Protein Structure
PSA	Polar Surface Area
P_{SB}	Structure-based probability
PTR1	Pteridine Reductase-1
Q2	Cross-validation coefficient
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationship
Q-SiteFinder	A program for binding site identification in proteins
R²	Determination coefficient
ReLeaSE	Reinforcement Learning for Structural Evolution
Relibase	A database for ligand-receptor interactions
RF	Random Forest

RF Algorithm	Random Forest Algorithm
RL	Reinforcement Learning
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
RoG	Radius of Gyration
SAR	Structure-Activity Relationship
SASA	Solvent-Accessible Surface Area
SB	Structure-based probability
SBDD	Structure-Based Drug Design
SbV	Pentavalent Antimonial Compounds.
SDF	Special Data File
SI	Selectivity Index
SIMCA-P+	Soft Independent Modeling of Class Analogy - Plus
SL	Sesquiterpene Lactone
SMILES	Simplified Molecular Input Line Entry System
SURFNET	Surface Identification of Ligand Binding Sites
SVM	Support Vector Machine
TES	N-[tris(hydroxymethyl)-methyl]-2-aminoethanesulfonic acid
THB	Tetrahydrobiopterin
THF	5,6,7,8-Tetrahydrofolate
TN	True-negative rate
TNBC	Triple-negative breast cancer
TP	True-positive rate
TPSA	Topological Polar Surface Area
TR	Trypanothione Reductase
TRAPP	Transient Pockets in Proteins
TS	Trypanothione Synthetase
vdW	van der Waals
VERIFY 3D	A program for verifying the compatibility of a protein 3D model
VL	Visceral Leishmaniasis
Vo	Initial velocity rate

VS	Virtual Screening
WHAT IF	Web-based Interface to GROMACS trajectory analysis tools
WHO	World Health Organization
X-ray	X-ray Crystallography
ZINC	A free database for virtual screening
μM	Micromolar

SUMÁRIO

1. INTRODUÇÃO	21
2. OBJETIVOS	27
2.1 Objetivos gerais.....	27
2.2 Objetivos específicos	27
 CAPÍTULO I	29
Drug discovery paradigms: target-based drug discovery.....	32
CAPÍTULO II	64
Discovery of Alternative Chemotherapy Options for Leishmaniasis via Computational Studies of Asteraceae.....	67
CAPÍTULO III	97
Selection of Antileishmanial Sesquiterpene Lactones from SistemX Database using a combined ligand/structure-based virtual screening approach.....	100
CAPÍTULO IV	133
Parte A.....	136
Identification of Kaurane-Type Diterpenes as Inhibitors of <i>Leishmania</i> Pteridine Reductase I.....	137
Parte B.....	172
Kaurane-Type Diterpenoids as Potential Inhibitors of Dihydrofolate Reductase- Thymidylate Synthase in New World <i>Leishmania</i> Species.....	173
CAPÍTULO V	195
Machine-Learning- and Structure-Based Virtual Screening for Selecting Cinnamic Acid Derivatives as <i>Leishmania major</i> DHFR-TS Inhibitors	199
 CONCLUSÃO	238
 ANEXOS I: Produção científica	241
ANEXOS II: Eventos e Congressos	250

The background of the page is a light gray abstract design. It features a complex network of thin lines connecting small dots, resembling a molecular structure or a data network. Interspersed among these lines are several hexagonal shapes of varying sizes and shades of gray, some solid and some outlined. The overall effect is a subtle, scientific, and technological aesthetic.

INTRODUÇÃO

1. Introdução

As doenças tropicais negligenciadas (NTDs) constituem um grupo de 20 condições que predominam em áreas tropicais, afetando principalmente populações empobrecidas e tendo um impacto desproporcional em mulheres e crianças [1]. A incidência global padronizada por idade e o número de casos de NTDs aumentaram entre 1990 e 2019 [2]. Para o período de 2020-2021, foi relatado que 1,65 bilhão de pessoas precisavam de tratamento em massa ou individual e cuidados para as NTDs, experimentando uma leve redução diretamente relacionada à pandemia de Covid-19. Essa emergência alterou tanto as cifras quanto o monitoramento desse grupo de doenças. Por exemplo, no Brasil, observou-se uma queda significativa nas consultas de casos de leishmaniose visceral (LV) em 2020 [3]. No entanto, o percentual da população mundial que poderia ser afetado por essas doenças permanece alto, gerando consequências sociais e econômicas para cerca de 20% da população global, indicando que as NTDs continuam sendo um significativo problema de saúde pública global [4,5].

Um dos principais desafios para o controle e eliminação desse grupo de doenças é a falta de atenção da maioria das empresas farmacêuticas, pois a pesquisa e produção de novos medicamentos para o tratamento delas não representariam um lucro significativo [6]. Os tratamentos atuais têm múltiplas limitações, como baixa eficácia, alta toxicidade e durações prolongadas. Desde a década de 1950, compostos antimoniais pentavalentes têm sido usados como tratamento principal para leishmanioses visceral (LV) e cutânea (LC), pois apresentam um melhor índice terapêutico do que medicamentos de segunda linha como anfotericina B, pentamidina e outros. Apesar de sua alta eficácia, os compostos antimoniais pentavalentes muitas vezes estão associados a altas frequências de efeitos adversos, como dor musculoesquelética, distúrbios gastrointestinais, dor de cabeça e anorexia, além de efeitos graves como toxicidade cardíaca, hepática e pancreática [7,8].

Portanto, o desenvolvimento de novos medicamentos contra esse grupo de doenças é crucial, sendo os produtos naturais uma das melhores alternativas para buscar estruturas potencialmente ativas contra essas NTDs. O número de medicamentos derivados de produtos naturais (PNs) no total de lançamentos de medicamentos no mercado ao longo de quatro décadas representa uma fonte significativa de novas entidades farmacológicas. Atualmente, mais de 80% dos novos medicamentos

aprovados pela FDA (Food and Drug Administration) estão diretamente ou indiretamente relacionados a produtos naturais [10].

Uma das famílias mais interessantes a ser estudada nesse campo, devido à sua diversidade, é Asteraceae (Compositae). Muitas plantas na família Asteraceae têm importância econômica, medicinal e ornamental [11]. É a maior família de plantas com flores, com mais de 1.690 gêneros e 32.000 espécies relatadas em todo o mundo (principalmente nas Américas) [12, 13]. As atividades parasitárias de algumas espécies de Asteraceae foram demonstradas anteriormente, com a artemisinina, uma lactona sesquiterpênica de *Artemisia annua*, sendo notável. Foi aprovada para o tratamento da malária devido às suas propriedades de endoperoxídeo, e seu descobridor recebeu o Prêmio Nobel de Fisiologia ou Medicina em 2015 [14, 15].

Várias espécies de Asteraceae demonstraram forte atividade antioxidante, anti-inflamatória e antimicrobiana, além de propriedades diuréticas e cicatrizantes. Seus efeitos farmacológicos podem ser atribuídos a uma variedade de compostos fitoquímicos, incluindo polifenóis, ácidos fenólicos, flavonoides, acetilenos e triterpenos [16]. Cerca de 5.500 compostos lactonas sesquiterpênicas (SLs), diversificados em 40 tipos estruturais, predominantemente exibindo esqueletos básicos de germacranolídeo, eudesmanolídeo e guaianolídeo, foram identificados [17].

De 2000 a 2023, cerca de 9.000 espécies da família Asteraceae foram descobertas [12, 13]. Com esse crescimento na quantidade de dados sobre produtos naturais, que geralmente são registrados separadamente em inúmeros artigos e livros, a seleção de informações para pesquisa muitas vezes é desafiadora. Portanto, um procedimento adequado para armazenar e sistematicamente organizar informações de dados de metabólitos secundários recuperados é essencial, com bancos de dados emergindo como ferramentas-chave em estudos de química medicinal [18].

Dois exemplos-chave contendo informações sobre Asteraceae foram desenvolvidos no Brasil: o software Sistemax (<http://sistemax.ufpb.br>), um exemplo de banco de dados de produtos naturais desenvolvido pela Universidade Federal da Paraíba para fornecer informações para estudos quimiossistemáticos, desreplicação e correlações botânicas. Tem sido utilizado com sucesso em vários estudos *in silico* [18]. Por outro lado, o AsterDB (<http://www.asterbiochem.org/asterdb>) e o banco de dados interno Aster-BioChem, que contém centenas de estruturas químicas relatadas apenas em espécies de Asteraceae. O AsterDB é o primeiro banco de dados a fornecer acesso gratuito e é dedicado especificamente a essa família botânica [19].

Esse fortalecimento de bancos de dados nas últimas duas décadas permitiu que estudos computacionais emergissem como uma boa alternativa em química medicinal para ajudar a estabelecer o design experimental, encontrar novos medicamentos e comparar a estrutura molecular e a atividade por meio de uma das ferramentas mais importantes nessa área, a relação quantitativa estrutura-atividade (QSAR), gerando resultados bem-sucedidos a baixo custo em curto espaço de tempo [20, 21].

O design de medicamentos assistido por computador (CADD) oferece a possibilidade de projetar novos medicamentos *in silico*, o que, combinado com estratégias convencionais em laboratório, permite que os pesquisadores desenvolvam tratamentos eficazes em tempos mais curtos e a custos mais baixos. Especificamente, modelos de relação quantitativa estrutura-atividade (QSAR) podem ser usados para identificar padrões moleculares que podem ser modificados em um motivo molecular para maximizar a atividade. A metodologia de ancoragem molecular também pode ser usada para explorar os comportamentos de pequenas moléculas no local de ligação de uma proteína alvo [19].

Neste trabalho, serão realizados diversos estudos de quimioinformática da família Asteraceae, incluindo triagem virtual baseada em estrutura de ligante e receptor, modelos QSAR, simulações de dinâmica molecular, entre outros. Esses estudos visam identificar estruturas potencialmente ativas contra leishmanioses, usando três diferentes bancos de dados de metabólitos secundários presentes em Asteraceae, divididos em três grandes classes: lactonas sesquiterpênicas, diterpenos do tipo caurano e estruturas derivadas do ácido cinâmico, em busca de moléculas com potencial atividade multialvo. Além disso, as moléculas selecionadas com potencial atividade serão testadas *in vitro* para validar todos os cálculos computacionais realizados.

REFERÊNCIAS

- [1]. Leishmaniasis. Geneva: World Health Organization; 2021 (<https://www.who.int/health-topics/leishmaniasis>, accessed 15 December 2023).
- [2]. WANG, Wei; LI, Jian; HABIB, Mohamed R. Innovative tools to support the elimination of neglected tropical diseases (NTDs). **Frontiers in Microbiology**, v. 14, p. 1208113, 2023.
- [3]. PAUL, Anindita; SINGH, Sushma. Visceral leishmaniasis in the COVID-19 pandemic era. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 117, n. 2, p. 67-71, 2023.
- [4]. WORLD HEALTH ORGANIZATION et al. **Global report on neglected tropical diseases 2023**. World Health Organization.
- [5]. Safety in administering medicines for neglected infectious diseases. Geneva: World Health Organization; 2022 (<https://apps.who.int/iris/handle/10665/344059>, accessed 10 January 2023).
- [6]. BURROWS, Jeremy N. et al. The role of modern drug discovery in the fight against neglected and tropical diseases. **MedChemComm**, v. 5, n. 6, p. 688-700, 2014.
- [7]. HERRERA ACEVEDO, Chonny et al. Computer-aided drug design using sesquiterpene lactones as sources of new structures with potential activity against infectious neglected diseases. **Molecules**, v. 22, n. 1, p. 79, 2017.
- [8]. GONÇALVES-OLIVEIRA, Luiz Filipe et al. Assessing nystatin cream treatment efficacy against *Leishmania (L.) amazonensis* infection in BALB/c model. **Experimental Parasitology**, v. 250, p. 108547, 2023.
- [9]. BARAZORDA-CCAHUANA, Haruna Luz et al. Computer-aided drug design approaches applied to screen natural product's structural analogs targeting arginase in *Leishmania* spp. **F1000Research**, v. 12, 2023.
- [10]. NEWMAN, David J.; CRAGG, Gordon M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. **Journal of natural products**, v. 83, n. 3, p. 770-803, 2020.
- [11]. YIEN, Raíssa Mara Kao et al. Nanotechnology Promoting the Development of Products from the Biodiversity of the Asteraceae Family. **Nutrients**, v. 15, n. 7, p. 1610, 2023.

- [12]. BOHM, Bruce A.; STUESSY, Tod F. **Flavonoids of the sunflower family (Asteraceae)**. Springer Science & Business Media, 2001.
- [13]. MOHANTA, Yugal Kishore et al. Potential use of the Asteraceae family as a cure for diabetes: A review of ethnopharmacology to modern day drug and nutraceuticals developments. **Frontiers in Pharmacology**, v. 14, 2023.
- [14]. TU, Youyou. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. **Nature medicine**, v. 17, n. 10, p. 1217-1220, 2011.
- [15]. CHEONG, Dorothy HJ et al. Anti-malarial drug, artemisinin and its derivatives for the treatment of respiratory diseases. **Pharmacological research**, v. 158, p. 104901, 2020.
- [16]. ROLNIK, Agata; OLAS, Beata. The plants of the Asteraceae family as agents in the protection of human health. **International journal of molecular sciences**, v. 22, n. 6, p. 3009, 2021.
- [17]. BAINS, Savita et al. Production of Sesquiterpene Lactones in Family Asteraceae: Structure, Biosynthesis, and Applications. In: **Biomedical Research, Medicine, and Disease**. CRC Press, 2023. p. 589-604.
- [18]. SCOTTI, Marcus Tullius et al. SistemX, an online web-based cheminformatics tool for data management of secondary metabolites. **Molecules**, v. 23, n. 1, p. 103, 2018.
- [19]. HERRERA-ACEVEDO, Chonny et al. Discovery of alternative chemotherapy options for Leishmaniasis through computational studies of Asteraceae. **ChemMedChem**, v. 16, n. 8, p. 1234-1245, 2021.
- [20]. SOROKINA, M.; STEINBECK, C. Review on natural products databases: where to find data in 2020. **Journal of cheminformatics**, v. 12, p. 1-51, 2020.
- [21]. ROY, Kunal (Ed.). **Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development**. Elsevier, 2023.

The background of the slide is a light blue and white abstract pattern. It features a complex network of interconnected nodes and lines, resembling a molecular structure or a data network. There are also several hexagonal shapes scattered throughout, some of which are solid blue and others are white with blue outlines. The overall effect is a clean, modern, and scientific aesthetic.

OBJETIVOS

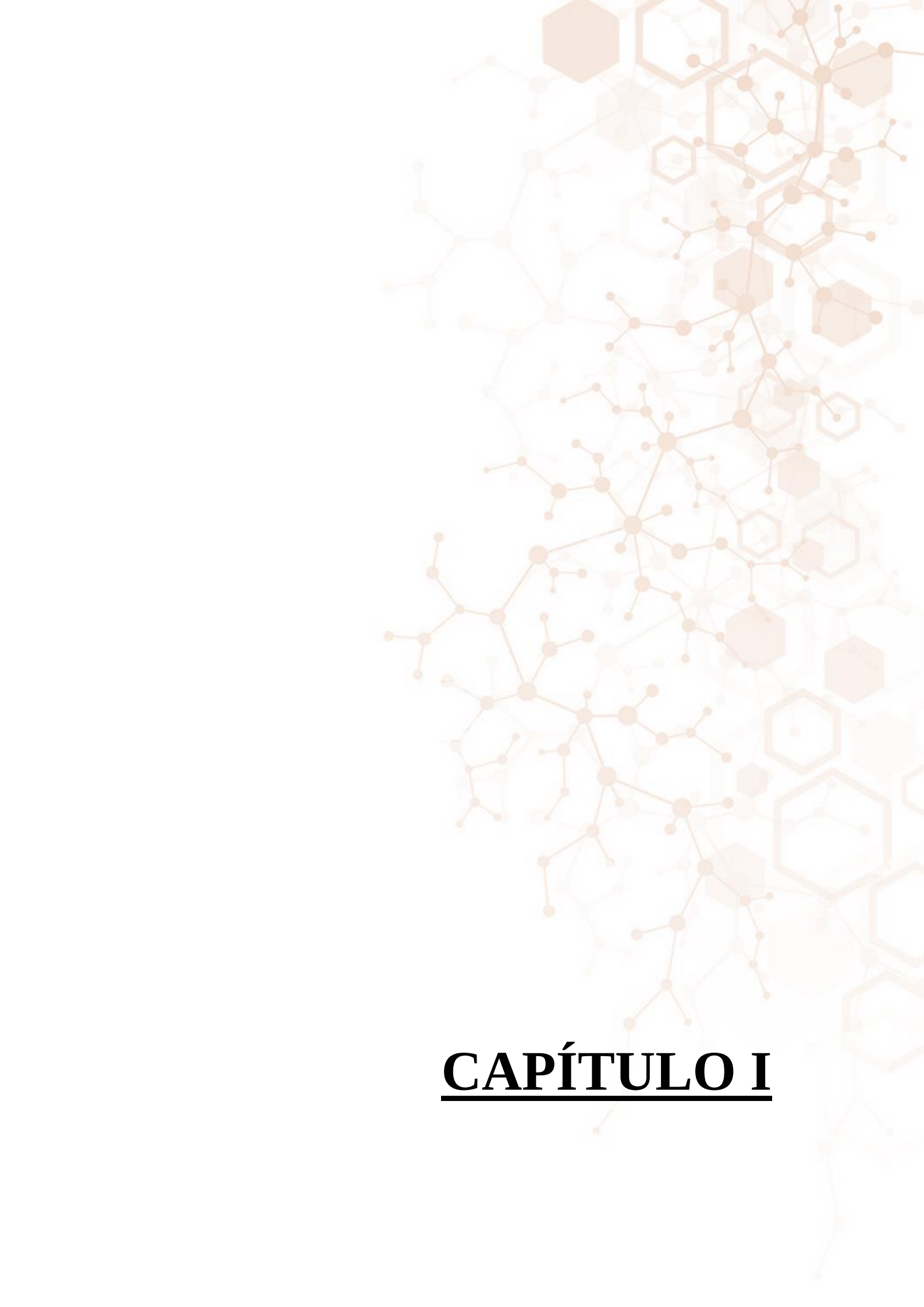
2. Objetivos

2.1 Objetivo geral

Realizar estudos quimioinformáticos da família Asteraceae para encontrar novas estruturas com um potencial terapêutico contra Leishmanioses, uma das principais doenças parasitárias negligenciadas.

2.2 Objetivos específicos.

- Construir bancos de metabólitos secundários da família Asteraceae usando o software SistematX.
- Criar modelos de relação estrutura-atividade quantitativos (QSAR) para prever a atividade de metabólitos secundários da família Asteraceae.
- Selecionar enzimas alvo envolvidas no metabolismo dos parasitas do gênero *Leishmania*, entendendo o mecanismo de ação de essas proteínas.
- Combinar as diversas aproximações de triagem virtual para a seleção de moléculas com potencial atividade contra as principais doenças parasitárias negligenciadas.
- Avaliar termodinamicamente os complexos entre enzimas do parasita e as estruturas selecionadas mediante simulações de dinâmica molecular.
- Verificar mediante ensaios *in vitro*, a atividade inibitória dos *hits* selecionados nos estudos computacionais.



CAPÍTULO I

O planejamento de fármacos auxiliado por computador, CADD (*Computer-Aided Drug Design*), tem sido significativo nos últimos anos para identificar moléculas com potencial atividade contra diversas doenças, incluindo as denominadas doenças tropicais negligenciadas, NTDs (*Neglected Tropical Diseases*). Esses métodos *in silico* tornaram-se uma alternativa viável, auxiliando na formulação de projetos experimentais, na descoberta de novos medicamentos e na análise comparativa de estruturas e atividades moleculares [1].

Dentro das metodologias de CADD, os métodos de desenho de medicamentos baseados em estrutura, SBDD (*Structure-based Drug Design*), incluindo docking molecular, simulações de dinâmica molecular e desenho *De novo*, estão entre os mais amplamente utilizados para a seleção de moléculas promissoras. Houve casos bem-sucedidos que resultaram no desenvolvimento de medicamentos disponíveis comercialmente, incluindo Amprenavir, Raltitrexed e Zanamivir [2].

Essas técnicas oferecem diversas vantagens dentro das ferramentas quimioinformáticas, como a redução no tempo e custo envolvidos na triagem de milhões de pequenas moléculas. Além disso, a avaliação computacional pode ser realizada antes da síntese física da molécula, e há uma variedade de ferramentas disponíveis para auxiliar em SBDD, tanto licenciadas quanto de acesso livre, facilitando seu uso [3]

No entanto, assim como todas as metodologias computacionais, algumas limitações estão associadas ao uso dessas ferramentas. Primordialmente, a especificidade para seu uso e a dificuldade em prever com precisão a posição correta de ligação e classificação de compostos devido à complexidade das interações de ligante-receptor são fatores a serem considerados para obter resultados confiáveis.

Devido à importância desse tipo de ferramentas na seleção de potenciais moléculas bioativas, no Capítulo 1 deste trabalho, foi realizada uma revisão abrangente das principais técnicas de SBDD, com foco nos fundamentos dessas técnicas, algoritmos de amostragem e pontuação, algumas vantagens e desvantagens associadas a essas metodologias, com ênfase em dois aspectos-chave: a identificação do alvo e do local de ligação.

REFERÊNCIAS

- [1]. HERRERA ACEVEDO, Chonny et al. Computer-aided drug design using sesquiterpene lactones as sources of new structures with potential activity against infectious neglected diseases. **Molecules**, v. 22, n. 1, p. 79, 2017.
- [2]. BATOOL, Maria; AHMAD, Bilal; CHOI, Sangdun. A structure-based drug discovery paradigm. **International journal of molecular sciences**, v. 20, n. 11, p. 2783, 2019.
- [3]. MAIA, Eduardo Habib Bechelane et al. Structure-based virtual screening: from classical to artificial intelligence. **Frontiers in chemistry**, v. 8, p. 343, 2020.

Drug discovery paradigms: target-based drug discovery

Chonny Herrera-Acevedo¹, Camilo Perdomo-Madriral², José Alixandre de Sousa Luis³, Luciana Scotti¹, Marcus Tullius Scotti^{1,*}.

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil.

² School of Science, Universidad de Ciencias Aplicadas y Ambientales, Calle 222 # 55–37, Bogotá D.C., Colombia.

³ Education and Health Center, Federal University of Campina Grande, Cuité 58175-000, PB, Brazil.

*Contact author. Email: mtscotti@gmail.com

Capítulo de livro publicado em “**Drug target Selection and Validation**”. Springer International Publishing. Part of the Computer-Aided Drug Discovery and Design book series (CADDD, volume 1).

DOI Number: https://doi.org/10.1007/978-3-030-95895-4_1

Ano de publicação: 2022

Abstract: Target-based drug discovery tools have been used with success in the pharmaceutical industry. They have become the fundamental methodologies for discovering new drugs in recent years, with two main advantages over the traditional methodologies: increased speed and greater economic efficiency. The improved computational capacities and new software packages have allowed the diversification and strengthening of these procedures. This chapter describes the main concepts related to target-based drug discovery, including two key steps, target and binding site identification, as well as the main features and limitations of the most common target-based methodologies: de novo drug discovery, molecular docking, and molecular dynamics.

Keywords: Target-based; molecular docking; molecular dynamics; drug discovery; sampling algorithms; scoring functions; De Novo drug discovery; binding site.

1. Introduction

Computer-aided drug design (CADD) has become the fundamental approach for the discovery, development, and analysis of structures with potential activity against many diseases in recent years. Two main types of approaches in CADD have been reported: ligand-based virtual screening (LBSS) and structure-based drug design (SBDD). These two approaches offer versatility and synergy, both in academia and industry [1]. SBDD, also known as target-based drug discovery, has as its main characteristic the use of computational methods and the three-dimensional (3D) structural information of the protein target to investigate the underlying molecular interactions involved in ligand–protein binding and thus interpret experimental results at an atomic level of detail [2].

Advances in structural resolution techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM) have allowed the development and strengthening of SBDD techniques [3]. Currently, the number of macromolecular structures registered in the Protein Data Bank (PDB) is close to 180,000 [4], significantly increasing the structural information about key macromolecular drug targets [1]. This evolution also drove the development of sophisticated software packages, facilitating *in silico* calculations of inhibitors into their predicted binding sites as well as the computational analysis of inhibitor binding and information on further enhancements [5].

Various examples of commercially available drugs developed from target-based drug design studies can be identified, including Amprenavir (Figure 1), which acts against two targets of human immunodeficiency virus (HIV), thymidylate synthase. Molecular dynamics calculations were used to explain the experimental observation that the P1' amide NH of substrate sequences was not required for binding and productive catalysis. From these results, and supported by *in silico* modeling, the N,N-dialkyl sulfonamide moiety was chosen to bind to the flap water molecule and to act as a scaffold for the P1' and P2' groups [6].

Other successful uses of SBDD include supporting the development of new drugs include Raltitrexed and Zanamivir (Figure 1). Raltitrexed acts against thymidylate synthase of the HIV [7]. Zanamivir was the first neuraminidase inhibitor to be marketed for the treatment of influenza. Using the GRID program, it was predicted that in the active site of the target, replacing the hydroxyl group at the 4-position of the ring by an

amine will improve the interactions with two neighboring glutamic acid residues, which identified this potent inhibitor of the neuraminidase enzyme ($K_i = 0.2 \text{ nM}$) [6].

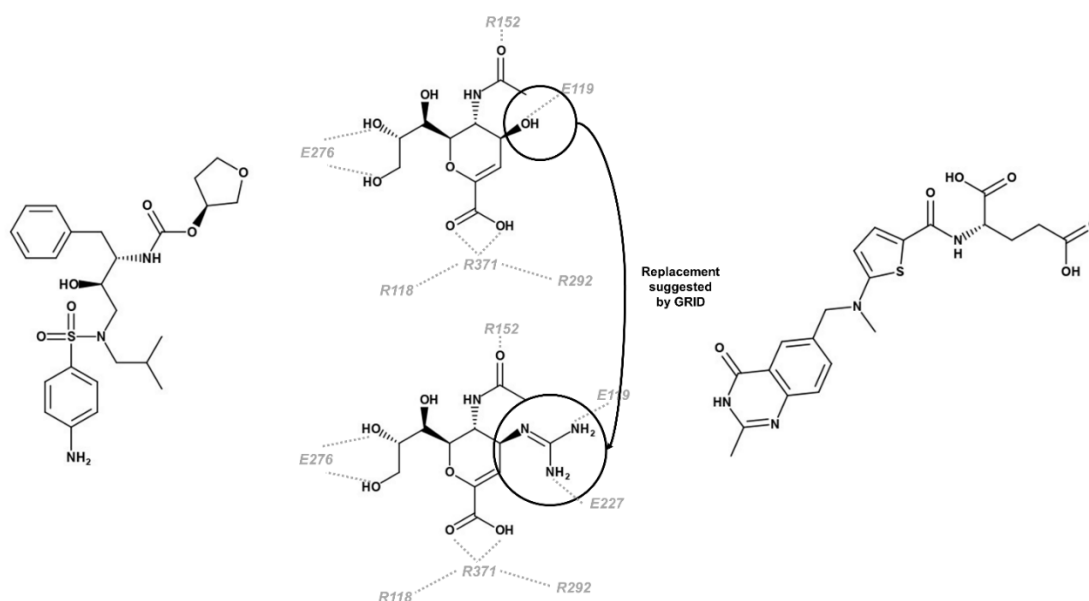


Figure 1. Examples of drugs commercially available that were developed from target-based drug design. *Left:* Amprenavir. *Center:* Zanamivir - GRID program predicted that the replacement of the hydroxyl group at the 4-position of the ring by an amine, would be improved the interaction with two neighboring glutamic acid (E) residues of the target (neuraminidase) [6]. *Right:* Raltitrexed.

In summary, the workflow of the SBDD process begins with target identification, which is supported by genetics, molecular biology, and bioinformatics methodologies. Next, protein extraction and purification are carried out. A structural determination of the target is performed, mainly using NMR, X-ray crystallography, and Cry-EM; for those proteins whose crystal structure is not defined, homology models are built in specialized software. Then, the biological assay is performed through different methodologies (Figure 2). The main three SBDD approaches are molecular docking, molecular dynamics, and de novo drug design. The small molecules to be evaluated in structure-based virtual screening (VS) methodologies are typically selected from databases of active compounds. Finally, the top hits are synthesized, and in-vitro tests are performed to identify the best structures.[7, 8]

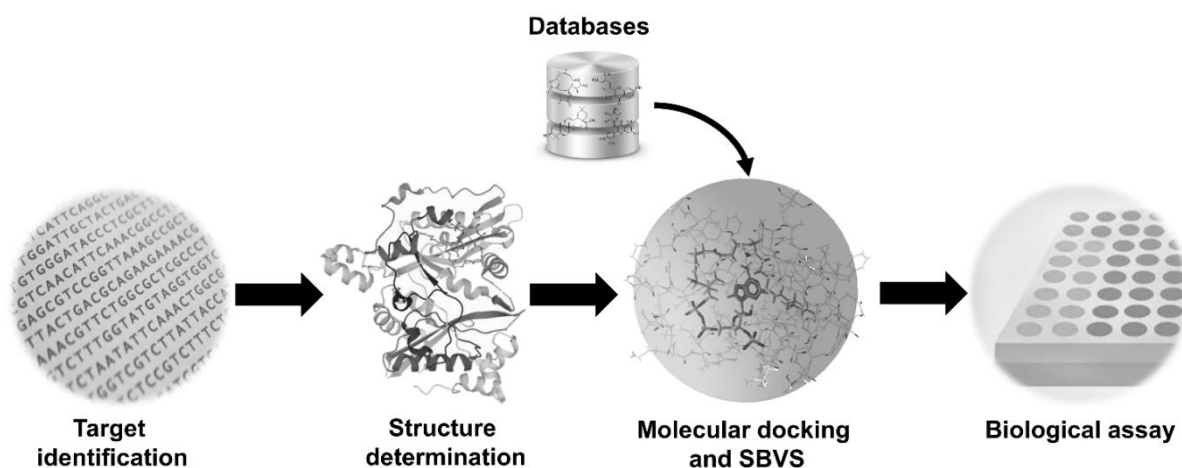


Figure 2. General workflow of the structure-based drug design (SBDD) [8].

This chapter reviews the main concepts in target-based drug design and describes the two key steps in these studies: identifying the target and the binding site. Also, we summarize the three main SBDD methodologies: molecular docking, molecular dynamics, and de novo drug design, discussing the basic concepts and classifications, as well as the algorithms and functions using for the calculations and the limitations of these methods.

2. Target Identification

When a research group proposes developing a new drug for a given disease, the first step is to understand the physiological and pathological processes of the disease. This study completed, it is possible to visualize possible molecular targets that are components of the human body or a particular pathogen. The main targets are proteins, such as enzymes, metabotropic or ionotropic receptors, antibodies, and nucleic acids, including nuclear and mitochondrial DNA in addition to messenger RNA. Protein targets are the most abundant since they mediate several important metabolic processes [9-11].

Drugs that act on enzymatic targets are mostly inhibitors; however, some can act as allosteric activators. In contrast, the drugs that act on membrane receptors can be agonists (when mimicking the effect of the endogenous substance) or antagonists (when blocking the signal promoted by the endogenous substance). That is, the mode of interaction determines the type of signal that will be transmitted to effectors located on the cytosolic side of the membrane. These effectors can be enzymes, ion channels, or sites for binding intracellular proteins. Examples of signal transduction pathways are

those in which the receptors are coupled to G proteins that in turn activate other enzymes that catalyze the conversion of triphosphate nucleotides into their cyclic variants that will act as second messengers. In other ways, these messengers can activate other enzymes and ion channels. Sometimes the channel itself can be the target, either through an allosteric receptor site located on the same protein or through direct interaction with the ion transport pore, which usually involves blocking the ion channel [12, 13].

Drugs that interact with nucleic acids can be obtained in two ways: i) by planning molecules that have the ability for direct interaction, that is, by forming covalent bonds or cleaving bonds, as well as by intercalation; ii) indirectly, through proteins that regulate gene replication and expression. These pathways are important for designing drugs to combat diseases caused by disorders in regulating the cell cycle, such as cancer [7, 11].

For the development of the SBDD, it is necessary to know the topological arrangement of the molecular targets and, for that, detailed 3D data of the macromolecule, obtained through X-ray crystallography techniques, NMR spectroscopy, and cryo-EM, is used. Such structures, when determined, are deposited in public databases that can be accessed freely. The most popular and widespread such databases are the PDB, InterPro, ExPASy, and Relibase [7, 10].

It is also possible to perform SBDD when the specific structure of a given macromolecule is not known using homology modeling. This method uses *in silico* methods, where it is possible to model a 3D structure from a homologous protein with a known structure. The three most well-known methods of predicting structures are comparative modeling, threading, and ab initio modeling, which are reliable and possible to validate. The first step in this process is to determine the registration or alignment of the target sequence that allows insertions and deletions in the experimental structure, which can be accomplished with several dynamic programming algorithms.

In the second stage, mutations are made to the amino acid residues of the experimental structure so that they correspond to those of the target protein. This strategy can also be used when more than one experimental structure is available, allowing the creation of a model with a hybrid structure that is closer to the structure of the target protein. The final step is to refine and examine the structure to ensure that it is reasonable and validates the model [14-16].

In the process of refining the model, one of the steps is the alignment of the structures, where identities and similarities are verified, the conserved regions are observed, and common amino acid residues in the active site are identified. A detailed investigation of these similarities allows the planning of new drugs more likely to be effective against the pathology under study.

There are tools available for the validation stage, such as the Ramachandran plot, which is useful because it defines the residues found in the most energetically favorable and unfavorable regions and guides the evaluation of the quality of the theoretical or experimental models of the proteins. This graph represents all possible combinations of the dihedral angles Ψ versus ϕ for each amino acid, except for glycine, which has no side chain. For the model to be considered reliable, at least 90% of the amino acids in the regions favorable to the study must be similar [17, 18].

The next step of SBDD is identifying the binding site on the macromolecular target and identifying the types of interactions necessary for the micromolecule to interact effectively and trigger a useful biological response. In this process, information about the free energy of the complex can be obtained, and the types of interactions between the atoms can be observed, leading to the identification of the best pharmacophores for the new ligand.

These binding sites are determined experimentally by X-ray crystallography techniques, including co-crystallization. The data obtained from the co-crystallized structures provide robust information about the binding sites and are very useful in understanding the interaction between the micro- and macromolecules. In some situations, only the macromolecule structures are deposited in the databases, without the presence of a ligand. There is no information about the connection site in such cases, so servers and online tools must be used to identify the sites.

Some examples of servers that can assist are DoGSite Scorer [19], CASTp [20], NSiteMatch [21], Metapocket [22], DEPTH [23], LISE [24], and MSpocket [25]. After the connection site has been identified, the volume of the connection pocket can be assessed using the tools TRAPP [26] and POVME [27]. More specifically, the residues in the macromolecules that favor an optimal interaction to trigger a biological response are identified. Therefore, it is necessary to know the interaction energies, van der Waals (vdW) forces, to obtain an optimal mapping of the connection site.

There are several methods for this purpose, one of which is Q-SiteFinder [28], which calculates the vdW interaction energies with a methyl probe. This method allows

the retention and grouping of those sites with more favorable energies based on their total interaction energies. From this knowledge, simulating the binding of various library compounds or compounds designed de novo in the active site of the protein allows the potential identification of novel drug candidates [7].

3. Binding site identification

Once the target protein has been identified and its biological effect has been confirmed, the next challenge in any SBDD research is identifying and validating the sites in the structure of the protein that will be more likely to interact with ligands. These binding sites, or cavities, can be classified into two different categories. If the interaction between the ligand and the target protein occurs in the catalytic site of the protein, it is called an orthosteric site; if the interaction takes place in a different region where the interaction produces changes in the protein conformation and the structure of the catalytic site, it is called an allosteric site [29].

Regardless of the type, understanding the cavity structure and its interaction energies with ligands is of great importance to any SBDD process. Therefore, several strategies have been developed to identify these binding sites in the target protein. These involve methods and algorithms that make use of information such as similarities of residues between functionally related proteins (evolutionary methods), structural features of the protein (geometrical methods), and interaction energy with probe molecules (energy-based methods) [30, 31].

3.1. Evolutionary methods

Based on the idea that homologous proteins (proteins that have a common ancestry) have some conserved residues, it is possible to assume that important regions of those proteins, such as those involved in biological or enzymatic activity, have functionally relevant sequences of residues that are conserved among different organisms to avoid malfunction in their physiological processes [32]. According to this assumption, information related to the cavities in a protein can be used in an evolutionary method to find the binding sites of any homologous or functionally related protein by comparing the residue sequence of their primary structures. Table 1 lists different algorithms based on this method.

Table 1. Examples of binding site identification software classified according to the type of algorithm used for the prediction [31, 33].

Evolutionary Methods	Geometrical Methods	Energy-based Methods
ConSurf	LIGSITE	GRID
Rate4Site	CAST	DrugSite
GarLig	SURFNET	QsiteFinder
	PocketPicker	MCSS
	VOIDOO	
	KVFinder	

The main drawback of this type of algorithm is the necessity of prior knowledge and the availability of information related to the sequences of existing proteins, which is practically impossible considering the huge number and diversity of proteins that exist in different organisms and species. Additionally, a low degree of similarity between the sequences of the target protein and the reported protein will lead to deficient cavity prediction results, interfering with the success of the SBDD study [30, 32].

3.2. Geometrical methods

Cavities are usually conceived as regions of large surface area in the protein and displayed as hole-shaped concave sections in the structure of the protein [33]. Geometry-based methods use the 3D structure of proteins and locate binding sites by considering their size, shape, and chemical properties (e.g., ionizability, intermolecular interaction tendency, and covalent bonding sites) [34]. In these methods, the protein structure and surface are analyzed using probe objects, typically spheres, that fit in the protein cavities, depending on their size [31, 34]. Other geometric techniques involve tessellation and 3D gridding of the protein structure to identify voids of a significant size on the protein surface. Thus, geometry-based methods can be classified as grid system scanning, probe sphere filling, and alpha-shape modeling [35].

Geometrical methods are the most frequently used for cavity detection in different software and algorithms [31]. Their greatest advantage is that these methods depend only on the availability of the 3D structure of the protein, and no additional prior information is needed. Since these algorithms use crystal structures from databases like the PDB, the protein structure is treated as static, disregarding its intrinsic flexibility and plasticity [36]. The most significant disadvantages for this category of methods come

from: i) the freedom and randomness considering the tessellation and gridding parameters that are needed to probe the protein surface, and ii) the inaccuracies that may arise from protein orientation effects and conformational changes that may occur [33].

3.3. Energy-based methods

As stated previously, to study the interaction energy of the molecules and the target protein, it is essential to predict their bonding probability and, consequently, to predict if the ligand will be an eligible potential drug candidate. Hence, another family of methods has been proposed to identify binding sites in target proteins, which is related to evaluating the interaction energy between different regions of the protein and some small molecules.

Energy-based methods identify cavities using only energetic criteria. In these algorithms, the vdW interaction energy, the hydrogen bond energy, or the Lennard–Jones potentials of a probe (a small molecule or fragment, such as water or a methyl group) and the target protein are evaluated [28, 35, 36]. These methods are usually more computationally demanding, which is a considerable disadvantage because of the interest in reduced calculation times in SBDD research [36].

4.. Target-based methodologies

4.1. Molecular docking

Molecular docking is the most common target-based drug discovery methodology. More than 52,000 papers related to target-based drug discovery have been published recently; more than 50% of them have been published in the last five years (2017–2021), (Web of Science, Copyright Clarivate 202). Molecular docking techniques aim to predict the binding mode of a ligand that best matches a macromolecular partner (i.e., proteins). These methodologies aim to accurately predict the structure of a ligand within the constraints of a receptor binding site and correctly estimate the strength of binding [37]. In protein–ligand docking, the process consists of generating several possible conformations and orientations, also called poses, of the ligand within the protein binding site [38].

Identifying the most likely binding conformations requires two steps: first, the exploration of a large conformational space representing the various potential binding modes; second, the accurate prediction of the interaction energy associated with each of the predicted binding conformations. All docking programs perform these two steps

through a cyclical process, in which the ligand conformation is evaluated by specific scoring functions until the results converge to a solution having minimum energy [1].

There are three types of molecular docking techniques, categorized by the flexibility of the molecules involved in the molecular docking calculations:

- *Rigid docking*: Both the ligand and the protein are considered rigid entities, and only the three translational and three rotational degrees of freedom are considered during sampling [38].
- *Semi-flexible docking*: One of the molecules, the ligand, is flexible, while the protein is rigid. Thus, the conformational degrees of freedom of the ligand are sampled, in addition to the six translational and rotational degrees of freedom [38].
- *Flexible docking*: The protein is not a passive rigid entity during binding, and both the ligand and the protein are considered flexible counterparts [38].

A molecular docking program has two essential components: sampling the conformations of the ligand in the active site of the protein and ranking these conformations via a scoring function [39, 40]. Sampling describes the generation of putative ligand-binding orientations and conformations near a binding site of the protein. Sampling can be further divided into two aspects, ligand sampling, and protein flexibility. Scoring explains the prediction of the binding tightness for individual ligand orientations and conformations using a physical or empirical energy function [41]. The computational cost required in the docking calculations increases according to the number of degrees of freedom. For this reason, both sampling and scoring should be optimized to give a good balance between accuracy and speed [38].

4.1.1. Sampling algorithms

Sampling algorithms evaluate the capability of each docking program to predict the ligand-binding poses [42]. Sufficient sampling of ligand and protein states in docking is essential, as well as accurate evaluation of the binding energies of potential protein–ligand complexes. A key issue is whether the docking program samples the possible states sufficiently and how increased sampling relates to improved scoring and outcomes. This includes sampling the internal degrees of freedom within the ligand, as well as sampling the poses between the ligand and the protein receptor [43]. The sampling algorithms are classified into three main categories (Table 2) [42]:

- i) *Shape matching*: These approaches consider the geometrical overlap between the two molecules. These algorithms can predict docking conformations rapidly; however, their accuracy is lower [44].
- ii) *Systematic search*: These approaches explore all ligand degrees of freedom during the search [45].
- iii) *Stochastic search algorithms*: These approaches randomly change all the degrees of freedom of the ligand (translational, rotational, and conformational) at each step, generating very diverse solutions. However, multiple independent runs of the algorithm are required to maximize the probability of finding the global energy minimum [45].

Table 2. Summary of the most common sampling algorithms for molecular docking.

Algorithm	Description	Examples	Ref.
Shape matching	Consider the geometrical overlap between two molecules. being able to predict docking conformations in fast speed, however at lower accuracy rate	ZDOCK SYSDOCK	[44]
Systematic search			
Exhaustive search	Explore the values of each degree of freedom in a combinatorial manner, rotating all dihedral angles of the ligand according to a predetermined range of values and a set of initial restraints	Glide eHiTS	[45]
Fragmentation	The ligand is separated in smaller fragments, followed by the selection, and docking of a base fragment into the receptor binding site. The ligand is then reconstructed incrementally by covalently linking the other fragments to the base group	FlexX	[45]
Conformation ensemble	Rigidly docks a set of previously generated ligand conformations into the binding site.	DOCK 4.0 FLOG	[45]
Stochastic search			
Monte-Carlo	The method involves applying random cartesian moves to the system and accepting or rejecting the move based on a Boltzmann probability	LigandFit, rDock	[46]
Genetic algorithm	Evaluating the evolution of a population of possible solutions via genetic operators to a final population,	rDock AutoDock	[46]

	optimizing a predefined fitness function. Degrees of freedom are encoded into genes or binary strings and the collection of genes, or chromosome, being assigned a fitness based on a scoring function.	GOLD	
Tabu search methods	An iterative procedure designed to obtain solution of optimization problems. The technique is defined as a Meta-Heuristic methodology that can move from a solution to another being able to save in memory the already visited solutions	PSI-DOCK	[44]
Swarm optimization methods	In each iteration, a particle moves based on the knowledge of other particles and its own experience to speculate about the promising region to explore. One global best solution is kept updated by all particles and each individual particle also keeps record of its own best solution.	PLANTS PSO-VINA	[47]

4.1.2. Scoring algorithms

In the molecular docking calculations, the scoring functions are responsible for distinguishing the correct poses (binders) from the incorrect poses (inactive compounds) in a reasonable computation time. The scoring functions estimate the binding affinity between the protein and the ligand by adopting various assumptions and simplifications [39].

Two theoretical aspects of these functions dominate their operational performance. The first is the degree to which a scoring function has a global extremum within the ligand pose landscape at the proper location. The second is the degree to which the magnitude of the function at the extremum is accurate [48].

In the docking process, the scoring algorithms have three aims:

- i) *Pose prediction*: The scoring function should be able to distinguish the experimental binding modes from all other modes explored through the searching algorithm [46].
- ii) *Virtual screening*: The capacity to classify active and inactive ligands [45].
- iii) *Binding affinity estimation*: Prediction of the affinity constants and correctly rank several compounds according to their potency [45].

The scoring functions are commonly classified into three general groups: force fields (FF), knowledge-based and empirical. Recently, a fourth group has appeared and become relevant, machine learning-based functions (Table 3) [7].

- *Force field scoring functions*: These employ energy functions from classical molecular mechanics (MM), which is the binding free energy of protein–ligand complexes defined by the sum of the vdW forces and the electrostatic interactions. The solvation is determined as a distance-dependent dielectric function, and the nonpolar contributions are assumed to be proportional to the solvent-accessible surface area [49].
- *Knowledge-based scoring functions*: These are based on the inverse Boltzmann statistic principle, which assumes that the frequency of different pairs of atoms at different distances is related to the interaction of the two atoms and converts their frequency into the distance-dependent potential of the mean force. Knowledge-based scoring functions have great advantages of reduced computing cost and predictive accuracy relative to other types of scoring functions [50].
- *Empirical scoring functions*: These functions compute the fitness of a protein–ligand binding by summing up the contributions of a number of individual terms, each representing an important energetic factor in protein–ligand binding. Multivariate linear regression (MLR) or partial least-squares (PLS) analysis is used in these empirical functions to determine the different related factors that affect the final binding score [51].
- *Machine learning-based functions*: dynamic techniques for constructing and optimizing models to predict a binding pose and affinity [7]. These functions have emerged in the last few years as potential rescoring tools for structure-based VS. The machine learning-based approaches can implicitly learn the function form from the training data and use adjustable parameters to improve performance further, thus offering higher performance flexibility and greater convenience [52]. The most common machine learning algorithms used as scoring functions are the support, vector machine, random forest, neural network, and deep learning algorithms [50].
- *Consensus scoring*. None of the four types of scoring functions mentioned above have general applicability or are perfectly accurate. Consensus scoring is used to

improve the probability of finding correct solutions by combining the scores from multiple scoring functions, using the advantages of the scoring functions while reducing the limitations. Suitable selection of the individual scores is fundamental to the design of the consensus analysis [53].

Table 3. Examples of the scoring functions in molecular docking calculations [53-56].

Force field-based	Empirical	Knowledge-based	Machine learning-based
AutoDock	ICM-Score	ICM-PMF	RF-Score
DOCK/FF	SYBYL/F-Score	SYBYL/PMF-Score3	RF-IChem
SYBYL/G-Score	SYBYL/ChemScore	DrugScore	SVM-IChem
SYBYL/D-Score	LUDI		
	X-Score		

4.1.3. Molecular docking limitations

Although molecular docking calculations have been a fundamental tool in target-based drug discovery and many successful cases have been reported, several limitations are still present that mainly affect the accuracy of the calculations and their computational cost [57]. The ability to computationally predict the thermodynamics of these molecular recognition processes has been relatively poor until recently [58] because of a lack of confidence in the scoring functions used to provide accurate binding energies [1].

The majority of docking studies have been performed using rigid conformations of the protein due to the high computational cost added by increasing the flexibility of the macromolecules; therefore, although the development of computers with greater calculation power has improved performance in flexible systems, integration of conformational changes with the scoring functions is still a problem to be solved [59].

Additionally, in molecular docking calculations, the solvation phenomenon is not explicitly considered, impeding the estimation of the desolvation energies. Water molecules have an important role in the ligand–target binding process because they are necessary to correct the free energy of binding associated with the ligand displacement of water molecules. Although some docking software such as Autodock4 and GOLD have implemented some approaches, solvation is an important challenge to resolve in the currently available docking scoring functions [1, 59].

The algorithms based on quantum mechanics (QM) and mixed methodologies such as the semiempirical and QM/MM methods have emerged as alternatives for improving the accuracy of binding energy calculations. The development of graphics processing units (GPUs) has supported an increase in the use of these types of techniques [59]. In general terms, QM-based scoring functions are better able to predict ligand affinities than MM-based functions. However, this is not observed in all cases: the agreement between experimentally measured activities and calculated binding energies is highly dependent on the chemical series under study [60].

Finally, the limitations related to the X-ray crystallographic data, which is the basis of the molecular docking calculations, are very important. Uncertainties in the atomic model can have significant consequences when this model is used as the basis for manual design, docking, scoring, and VS efforts [61]. The redocking procedure, where a known ligand is docked to the “induced-fit” form of the target [62], is often the only method of validating the docking parameters used; it is important to develop additional validation procedures, especially those that include biological information, in order to decrease potential errors related to the X-ray crystallographic data and improve the robustness of the molecular docking calculations.

4.2. Molecular dynamics simulations

One of the main limitations of the molecular docking calculations is related to the flexibility of the targets. A high percentage of protein–ligand systems evaluated by docking require rigid conditions to test a large number of molecules quickly with low computational cost by VS. To evaluate the physical movements of these systems, molecular dynamics (MD) simulations are used. MD is an *in silico* technique, which aims to derive statements about the structural, dynamical, and thermodynamical properties of the molecular systems [63].

Leimkuhler and Matthews define the MD method as developing quantitative predictions of molecular size and shape, flexibilities, interactions with other molecules, behavior under pressure, and the relative frequency of one state or conformation compared to another [64]. Historically, MD was developed in the early 1950s and has evolved constantly since then. Rahman and Verlet [65, 66] refined the technique by implementing the method for all states of matter. Part of the evolution and relevance of these techniques is related to the foundation of powerful programs found by Martin Karplus, Michael Levitt, and Arie Warshel that were used to understand and predict

chemical processes. These three researchers were recognized with the Nobel Prize in Chemistry in 2013 for the development of multiscale models for complex chemical systems [67]. Currently, more than 20,000 papers related to MD are published annually, with close to 6% annual growth in the number of works from 2016 to 2020 (Web of Science, Copyright Clarivate 202).

4.2.1 Algorithms in molecular dynamics simulations.

For a MM system, MD consists of iterations of the instant forces present and the consequent movements of that system. The MM system is described as a set of particles that move in response to their interactions according to Newton's equations of motion, and the MD simulation computes the movements of atoms with time by integrating these equations, given below [38, 68]:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i}$$

where $F_i(t)$ is the force exerted on atom i at time t , $r_i(t)$ is the vector position of the atom i at time t , and m_i is the mass of the atom [38].

The essential function of the MD algorithms is to derive Newton's equations in a time differential (dt) for each atom of the system [38].

- *Position:* $r_i(t) = (x_i(t), y_i(t), z_i(t))$
- *Velocity:* $v_i(t) = \frac{dr_i(t)}{dt}$
- *Acceleration:* $a_i(t) = \frac{d^2 r_i(t)}{dt^2} = \frac{f_i(t)}{m_i}$
- *Force:* $f_i(t) = -\frac{dV(r(t))}{dr_i(t)}$
- *Potential Energy:* $V(r(t))$

Where, x_i , y_i and z_i are the coordinates of the i atom in the time (t).

The algorithms used in MD are classified into five types: integrators, short-range interaction, long-range interaction, parallel computing, and ab initio (Table 4).

Table 4. Main classification of the most common MD algorithms

Short-range interaction [69]	Large-range interaction [70]	Integrators [68]	Paralleling computing [71]	Ab initio [72]
Verlet Cell-linked List (VCL)	Ewald summation PME P3M MSM	SHAKE Beeman's Verlet-Stoermer Leapfrog	Point-centered domain decomposition	Car-Parrinello

Two types of algorithms depend on the type of nonbonded interactions evaluated, short-range and large-range interactions. The short-range algorithms are based on the Lennard–Jones potential that represents the van der Waals interactions. Meanwhile, the large-range algorithms are based on the treatment of the long-range electrostatic potentials [70]. In the ab initio algorithms, the interactions between ions and electrons are treated fully by QM, and the ions are moved using the classical Newton equations of motion. Carr–Parrinello is one of the most common ab initio algorithms, which was developed to calculate the ab initio forces on the ions and keep the electrons close to the Born–Oppenheimer surface while the atoms move [72]. In parallel computing, MD simulations are performed on parallel computers: the molecular system is divided into clusters assigned to individual processors [71]. Finally, integrators consist of different algorithms to integrate the equations of motion. Many of these are difference methods, in which the integration is partitioned into small steps, each separated by a defined period because the continuous potentials describing atomic interaction preclude an analytical solution. Among the most common MD integrator algorithms identified are Verlet, Leapfrog, and Beeman [68].

4.2.2. Force field in Molecular dynamics simulations.

As observed previously in Molecular docking calculations, MM techniques use a force field to describe the dependence of the energy on the atomic coordinates of the system [73]. In MD, the accuracy of the force field is critical to the validity and stability of the simulations of proteins and, in fact, all macromolecules [68]. The main FFs used in MD include MM2, MM3, MM4, CHARMM, AMBER, GROMOS, OPLS, and

COMPASS. Many versions of these force fields exist to perform MD simulations, such as GROMOS96, GROMOS45A3, and GROMOS53A5 [73].

The formal expression of a force field is divided into two groups of terms, bounded and nonbounded. The bonded interactions account for the stretching of bonds, the bending of valence angles, and the rotation of dihedral angles. The nonbonded interactions capture electrostatics, dispersion, and the Pauli exclusion forces [74].

$$E_{Total} = E_{Bonded} + E_{nonbonded} + E_{others}$$

$$E_{bonded} = \sum_{Bonds} K_b(b - b_o)^2 + \sum_{angles} K_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} K\chi[1 + \cos(n\chi - \sigma)]$$

$$E_{nonbonded} = \sum_{\substack{nonbonded \\ pairs\ ij}} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 * \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right)$$

The first part of the energy equation is related with the bonded interactions and has three terms. The first describes the stretching of bonds, where b is the interatomic distance (bond length) and K_b and b_0 are the parameters describing the stiffness and the equilibrium length of the bond, respectively. In agreement with the assumption in MM treating the bonded interactions as springs, this term also has a quadratic form known as Hooke's Law. For the calculation of the second term, the involvement of three atoms is necessary to describe the bending of angles, where θ is the angle formed by the two bond vectors; K_θ and θ_0 are the parameters describing the stiffness and equilibrium geometry of the angle, respectively. These parameters are similar to the terms for bond stretching, K_b and b_0 . The last term is related to the energy associated with the rotation of dihedral angles defined by four atoms, where χ is the dihedral value, $K\chi$ is the energetic parameter that determines the barrier heights, n is the periodicity or multiplicity, and σ is the phase [38, 74].

Nonbonded interactions have two terms. The first is known as the Lennard-Jones equation. The second term relates the models of attractive dispersion and repulsive Pauli exclusion interactions and is commonly referred to as the van der Waals term [38, 74]. Calculating these nonbonded interactions in biomolecular simulations is a key issue and one of the main challenges in the area [75].

Since the early 1980s, when the most common protein FFs, Amber, CHARMM, and OPLS, were developed, protein FFs have continuously evolved and improved [76].

Currently, protein FFs are highly advanced, even with respect to other macromolecules. Guvench and Mackerell performed a comparison among the most common protein FFs for MD and found that all the studied force fields (Amber, CHARMM, GROMOS, and OPLS-AA) treat proteins at an often satisfactory level of accuracy [74]. Developments in these methods have produced multi-microsecond simulations of two proteins, ubiquitin and Protein G, using a number of different FFs. Four FFs (CHARMM22*, CHARMM27, Amber ff99SB-ILDN, and Amber ff99SB*-ILDN) showed a good agreement between NMR data and MD simulations [77].

4.2.3. Limitations of Molecular dynamics simulations.

Gonzalez chose an interesting definition of the utility of the MD simulations: the aim is not to reproduce an experimental result but to understand the microscopic origin of the physical properties observed or to predict qualitatively the behavior expected at conditions that cannot be accessed experimentally [73]. In protein FFs, the constant evolution has allowed a high level of accuracy in the predictions. However, the same level of accuracy is not observed for other macromolecules since only a few specialized force fields are regularly used today for sugars, nucleic acids, and lipids [75]. For example, Ricci performed a comparison of FF and terminal nucleotide definitions because, despite the importance of DNA as a target for several proteins and drugs, molecular dynamics simulations with nucleic acids still face many challenges, such as the reliability of the chosen force fields [78].

The computational cost is one of the main limitations of MD simulations. The conformational sampling of biological systems is in many cases limited by the capabilities of the computational hardware [79]. Even though other techniques have been developed to overcome the limitations of MD, such as enhanced sampling MD simulations, in classical molecular dynamics, a low computational cost requires some approximations, which decrease the accuracy of the predictions. Ab initio methods produce more realistic simulations of complex molecular systems and processes but are computationally complex. To overcome this difficulty, electronic interactions are approximated with an effective pseudopotential, and the orbitals are expressed in terms of a suitable functional basis. Hybrid QM/MM methods follow a similar approach [80].

Moreover, the size of the biological system is a key factor in MD simulations. Some biomolecular processes, including ligand binding and conformational change, often take place on timescales longer than those accessible with a classical all-atom MD

simulation. For some systems with about 50,000 atoms, one GPU takes a few days to simulate a single microsecond [81]. Through Markov State Models (MSMs), these limitations have been overcome. MSMs are based on an ensemble view of the dynamics, from which statistical properties, such as the probability of a state being occupied and the probability of jumping from one state to another, are computed [38].

4.3. *De novo* drug design

The development of new molecular structures and entities with therapeutical uses is an enormous challenge in producing new drugs for commercial use. Consequently, different strategies have been proposed to find and design new candidates to fulfill this demand. Considering the gigantic amount of information available to explore the extant chemical space associated with different diseases, it is possible to approach the drug design task, where this information is valuable and convenient, by proposing new molecular entities from small molecular fragments assembled to maximize the interaction of the ligand with the active site, by means of using computational growth algorithms, through a method called *de novo* drug design (DNDD) [82].

De novo design can be defined as a method where a molecule is designed to satisfy the constraints and characteristics needed to achieve the required biological or therapeutical activity [83]. There are two useful ways to start a *de novo* methodology. First, in the structure-based approach, the construction of the molecules can be started using the knowledge acquired from the structure of the active site: the 3-D structure and the possible interaction sites are known. The second approach is the ligand-based approach, in which the construction of molecules is started from a known active ligand, but the active site information is unavailable and hard to obtain [82]. Once the methodology has been chosen, the next step is to build and evaluate the molecules that satisfy the established restrictions [82, 83]. Some popular software for these approaches is shown in Table 5 and classified according to the methodology used.

Table 5. Examples of *De novo* software classified according to the type of methodology used to construct and score molecules [82, 84].

Fragment-based Software		
	Ligand-based	Structure-based
LUDI		X
SPROUT		X
Chemical Genesis	X	X
PRO_LIGAND	X	X
TOPAS	X	
ADAPT		X
Atom-based Software		
	Ligand-based	Structure-based
Diamond Lattice		X
LEGEND		X
MCDNLG		X
DLD		X
RASSE		X

4.3.1. Sampling, Scoring and Optimization

To assemble a list of candidate molecules, there are two possible methodologies. First, in the atom-based methodology, the transformation of the molecules is achieved by modifying one atom into another to explore the molecular possibilities. This is a very intuitive method, which results in a wider ensemble of novel structures. However, it suffers from the possibility of creating new molecules with unfavorable structures, unstable hetero–hetero atomic bonds, or difficult synthetic accessibility [82, 85, 86]. Second, in the fragment-based approach, the design of the molecules is achieved by the mutation of pre-defined molecular fragments, which causes a vast reduction of the possibilities within the chemical space, is less time-consuming, and the chemical feasibility of the proposed molecules is higher (see Table 5) [85, 86].

During the construction of the molecules, they are scored according to a set of suitable molecular descriptors to determine their druggability using different algorithms. If the study is done using a structure-based methodology, the algorithm calculates the score based on the interaction of the fragments with the active site by docking procedures, and those fragments are then used as seeds to build the rest of the molecule.

In contrast, in ligand-based approaches, molecules or fragments are compared to a reference ligand by considering its similarity to their descriptors.

Optimization processes are typically based on evolutionary computation, which is inspired by biological evolution, involving mutation, crossover, and selection of the candidates more likely to “survive” based on their performance in the defined biological function. Evolutionary computation is divided into four categories: genetic algorithms, genetic programming, evolution strategies, and evolutionary programming [82, 87]. In general, all of these algorithms begin with a population of candidates, within which the most promising molecules are selected, finally proposing a new generation of candidates to replace the initial one, and then the algorithm restarts. The main advantages of these algorithms are their simplicity, adaptability, and efficacy in exploring the chemical space [87].

4.3.2. Machine Learning in *De novo* drug design

Regarding the vast amount of information available about existing compounds and their biological activity in databases such as ChEMBL, COCONUT [88], and ZINC, several methods have been developed to analyze and predict the chemical behavior of these bioactive compounds, causing a considerable reduction in time and allowing a more efficient exploration of the chemical space in *de novo* studies [89]. In particular, due to the advances achieved in machine learning, the possibility of allowing computers to learn to select a promising molecule by starting from chemical and biological information, as well as giving them the ability to find structural fingerprints that can lead to the identification of the most important characteristics of molecules to enhance their activity, has had a remarkable effect on *de novo* methodologies.

Deep Learning (DL) [90-92], Reinforcement Learning (RL) [93, 94], Deep Reinforcement Learning (DRL) [95] and Reinforcement Learning for Structural Evolution (ReLeaSE) [96] are some of the machine learning methods developed to assist with *de novo* design of molecules in the last decade [82, 96]. For all these methods, it is necessary to start from molecular information, usually in the SMILES format, and the biological activity measurements of the selected target. The main differences between these methods are the type of architecture developed for data analysis and evaluation and the evaluated molecular properties.

4.3.3. Limitations in *De novo* drug design

Despite all the advantages of *de novo* methodologies, especially the possibility of extensively exploring the chemical space for lead structure identification, some existing disadvantages and limitations exist. The most relevant limitation is related to the accuracy of the scoring functions. *De novo* software does not consider the conformational modifications that may take place in the structure of the target protein (particularly in the active site) or the entropic effects caused by the solvent–ligand interaction, which limits the calculation of the ligand–pocket interaction energy and thus reduces the accuracy of the molecular design [97].

Additionally, *de novo* designed molecules might have good predicted selectivities and interaction energies with targets; however, these methodologies do not allow the prediction of other important physicochemical properties of the designed structures (e.g., solubility, permeability to cells, and affinity to transport proteins) involved in pharmacokinetic studies and relevant to choosing the most promising therapeutical candidates [82, 84]. These limitations need to be overcome in order to increase the success of *de novo* methodologies in the design of effective drug molecules.

5. Conclusions and future directions

In this chapter, we reviewed the main aspects of target-based drug design to explain the development and importance of these types of techniques in drug discovery. Various structures have been successfully discovered through this class of methods, but it is still necessary to overcome some existing limitations of these methodologies.

The constantly growing and updated databases are the main source of structures for methodologies such as VS and online software for molecular docking calculations present a promising future of target-based drug design. However, for the continued growth of these databases to be useful, the correct use of these techniques is necessary, which begins with understanding the main concepts of the methodologies.

For the molecular docking calculations, the development of scoring functions is critical for flexible receptor docking to improve the energy prediction and the correct spatial position of the ligands. In the same way, it is fundamental to develop new scoring algorithms to improve the accuracy, such as those based on QM or mixed methodologies, including the semiempirical and QM/MM methods and the corresponding validation procedures of the calculations.

Moreover, the importance of MD simulations as a key technique in the drug development process is likely to grow substantially with increasing computer power and advances in the development of FFs and enhanced MD methodologies [63]. In addition, improving and developing the FF for macromolecules such as sugars, nucleic acids, and lipids is vital to expand the applications of the MD simulations in biological systems.

In all cases, the main challenges in target-based drug design methodologies are related to improving the accuracy of the predictions while reducing the computational cost of the calculations.

References

- [1].Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D., Molecular docking and structure-based drug design strategies. *Molecules* **2015**, 20, 13384-13421.
- [2].Lionta, E.; Spyrou, G.; K Vassilatis, D.; Cournia, Z., Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry* **2014**, 14, 1923-1938.
- [3].Wang, W., Applications of biophysical methods in small-molecule modulators targeting protein function. **2021**.
- [4].Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic acids research* **2000**, 28, 235-242.
- [5].Van Montfort, R. L. M.; Workman, P., Structure-based drug design: aiming for a perfect fit. *Essays in biochemistry* **2017**, 61, 431-437.
- [6].Clark, D. E., What has computer-aided molecular design ever done for drug discovery? *Expert opinion on drug discovery* **2006**, 1, 103-110.
- [7].Batool, M.; Ahmad, B.; Choi, S., A structure-based drug discovery paradigm. *International journal of molecular sciences* **2019**, 20, 2783.
- [8].Ramesh, S.; Vallinayagam, S.; Rajendran, K.; Rajendran, S.; Rathinam, V.; Ramesh, S. Computer-Aided Drug Designing–Modality of Diagnostic System. In *Biomedical Signal Processing for Healthcare Applications*; CRC Press, pp 195-218.
- [9].Lappano, R.; Maggiolini, M., G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature reviews Drug discovery* **2011**, 10, 47-60.
- [10].Valeur, E.; Guéret, S. M.; Adihou, H.; Gopalakrishnan, R.; Lemurell, M.; Waldmann, H.; Grossmann, T. N.; Plowright, A. T., New modalities for challenging

targets in drug discovery. *Angewandte Chemie International Edition* **2017**, 56, 10294-10323.

[11].Surabhi, S.; Singh, B. K., Computer aided drug design: an overview. *Journal of Drug Delivery and Therapeutics* **2018**, 8, 504-509.

[12].Bagal, S. K.; Brown, A. D.; Cox, P. J.; Omoto, K.; Owen, R. M.; Pryde, D. C.; Sidders, B.; Skerratt, S. E.; Stevens, E. B.; Storer, R. I., Ion channels as therapeutic targets: a drug discovery perspective. *Journal of medicinal chemistry* **2013**, 56, 593-624.

[13].Zhu, F.; Shi, Z.; Qin, C.; Tao, L.; Liu, X.; Xu, F.; Zhang, L.; Song, Y.; Liu, X.; Zhang, J., Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic acids research* **2012**, 40, D1128-D1136.

[14].Oda, A., Development and validation of programs for ligand-binding-pocket search. *Yakugaku zasshi: Journal of the Pharmaceutical Society of Japan* **2011**, 131, 1429-1435.

[15].Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N., Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Computational and structural biotechnology journal* **2013**, 5, e201302011.

[16].Wang, T.; Wu, M.-B.; Zhang, R.-H.; Chen, Z.-J.; Hua, C.; Lin, J.-P.; Yang, L.-R., Advances in computational structure-based drug design and application in drug discovery. *Current topics in medicinal chemistry* **2016**, 16, 901-916.

[17].Maia, M. d. S.; Nunes, T. A. d. L.; Sousa, J. M. S. d.; Rodrigues, G. C. S.; Monteiro, A. F. M.; Tavares, J. F.; Rodrigues, K. A. d. F.; Mendonça-Junior, F. J. B.; Scotti, L.; Scotti, M. T., Virtual screening and the in vitro assessment of the antileishmanial activity of lignans. *Molecules* **2020**, 25, 2281.

[18].Potapov, V.; Cohen, M.; Inbar, Y.; Schreiber, G., Protein structure modelling and evaluation based on a 4-distance description of side-chain interactions. *Bmc Bioinformatics* **2010**, 11, 1-17.

[19].Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M., DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **2012**, 28, 2074-2075.

[20].Sahu, A.; Patra, P. K.; Yadav, M. K.; Varma, M., Identification and characterization of ErbB4 kinase inhibitors for effective breast cancer therapy. *Journal of Receptors and Signal Transduction* **2017**, 37, 470-480.

- [21].Sun, J.; Chen, K., NSiteMatch: Prediction of binding sites of nucleotides by identifying the structure similarity of local surface patches. *Computational and mathematical methods in medicine* **2017**, 2017.
- [22].Huang, B., MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS A Journal of Integrative Biology* **2009**, 13, 325-330.
- [23].Tan, K. P.; Nguyen, T. B.; Patel, S.; Varadarajan, R.; Madhusudhan, M. S., Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic acids research* **2013**, 41, W314-W321.
- [24].Xie, Z.-R.; Liu, C.-K.; Hsiao, F.-C.; Yao, A.; Hwang, M.-J., LISE: a server using ligand-interacting and site-enriched protein triangles for prediction of ligand-binding sites. *Nucleic acids research* **2013**, 41, W292-W296.
- [25].Zhu, H.; Pisabarro, M. T., MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics* **2011**, 27, 351-358.
- [26].Stank, A.; Kokh, D. B.; Horn, M.; Sizikova, E.; Neil, R.; Panecka, J.; Richter, S.; Wade, R. C., TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. *Nucleic acids research* **2017**, 45, W325-W330.
- [27].Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E., POVME 3.0: software for mapping binding pocket flexibility. *Journal of chemical theory and computation* **2017**, 13, 4584-4592.
- [28].Laurie, A. T. R.; Jackson, R. M., Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics* **2005**, 21, 1908-1916.
- [29].Nussinov, R.; Tsai, C.-J., The different ways through which specificity works in orthosteric and allosteric drugs. *Current pharmaceutical design* **2012**, 18, 1311-1316.
- [30].Harigua-Souiai, E.; Cortes-Ciriano, I.; Desdouits, N.; Malliavin, T. E.; Guizani, I.; Nilges, M.; Blondel, A.; Bouvier, G., Identification of binding sites and favorable ligand binding moieties by virtual screening and self-organizing map analysis. *BMC bioinformatics* **2015**, 16, 1-15.
- [31].Oliveira, S. H. P.; Ferraz, F. A. N.; Honorato, R. V.; Xavier-Neto, J.; Sobreira, T. J. P.; de Oliveira, P. S. L., KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC bioinformatics* **2014**, 15, 1-8.
- [32].Armon, A.; Graur, D.; Ben-Tal, N., ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *Journal of molecular biology* **2001**, 307, 447-463.

- [33].Simões, T.; Lopes, D.; Dias, S.; Fernandes, F.; Pereira, J.; Jorge, J.; Bajaj, C.; Gomes, A. Geometric detection algorithms for cavities on protein surfaces in molecular graphics: a survey. 2017, Wiley Online Library; Vol. 36; pp 643-683.
- [34].Laskowski, R. A., SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of molecular graphics* **1995**, 13, 323-330.
- [35].Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M., Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* **2010**, 26, 46-52.
- [36].Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X., MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, 27, 3276-3285.
- [37].Yuriev, E.; Ramsland, P. A., Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition* **2013**, 26, 215-239.
- [38].Salmaso, V.; Moro, S., Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in pharmacology* **2018**, 9, 923.
- [39].Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M., Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* **2011**, 7, 146-157.
- [40].Tripathi, A.; Bankaitis, V. A., Molecular docking: From lock and key to combination lock. *Journal of molecular medicine and clinical applications* **2017**, 2.
- [41].Huang, S.-Y.; Zou, X., Advances and challenges in protein-ligand docking. *International journal of molecular sciences* **2010**, 11, 3016-3034.
- [42].Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T., Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics* **2016**, 18, 12964-12975.
- [43].Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K., Ligand pose and orientational sampling in molecular docking. *PloS one* **2013**, 8, e75992.
- [44].Dias, R.; de Azevedo, Jr.; Walter, F., Molecular docking algorithms. *Current drug targets* **2008**, 9, 1040-1047.
- [45].Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E., Receptor–ligand molecular docking. *Biophysical reviews* **2014**, 6, 75-87.

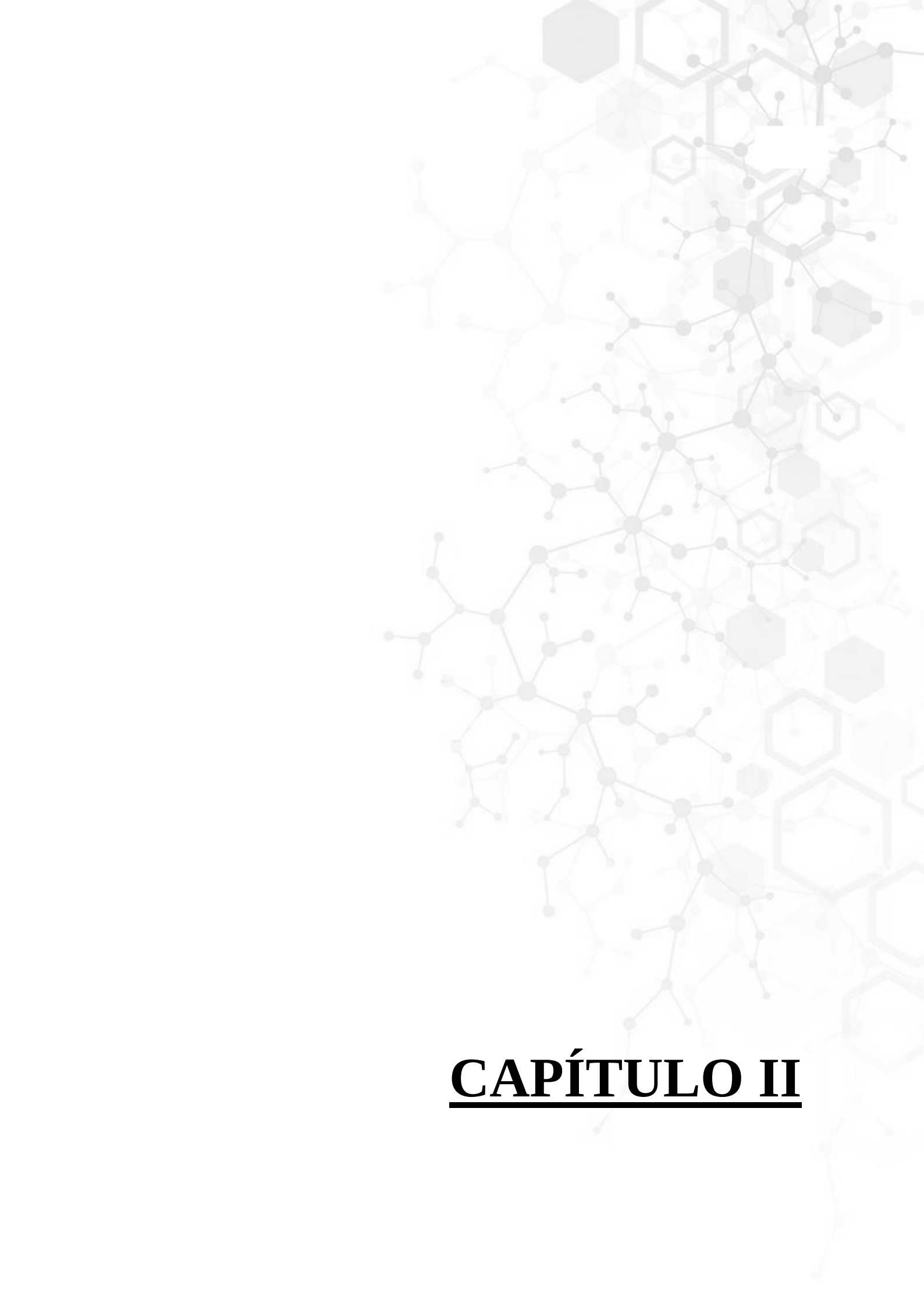
- [46].Taylor, R. D.; Jewsbury, P. J.; Essex, J. W., A review of protein-small molecule docking methods. *Journal of computer-aided molecular design* **2002**, 16, 151-166.
- [47].Ng, M. C. K.; Fong, S.; Siu, S. W. I., PSOVina: The hybrid particle swarm optimization algorithm for protein–ligand docking. *Journal of bioinformatics and computational biology* **2015**, 13, 1541007.
- [48].Jain, A. N., Scoring functions for protein-ligand docking. *Current Protein and Peptide Science* **2006**, 7, 407-420.
- [49].Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., Assessing scoring functions for protein– ligand interactions. *Journal of medicinal chemistry* **2004**, 47, 3032-3047.
- [50].Li, J.; Fu, A.; Zhang, L., An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences* **2019**, 11, 320-328.
- [51].Liu, J.; Wang, R., Classification of current scoring functions. *Journal of chemical information and modeling* **2015**, 55, 475-482.
- [52].Shen, C.; Hu, Y.; Wang, Z.; Zhang, X.; Pang, J.; Wang, G.; Zhong, H.; Xu, L.; Cao, D.; Hou, T., Beware of the generic machine learning-based scoring functions in structure-based virtual screening. *Briefings in Bioinformatics* **2021**, 22, bbaa070.
- [53].Huang, S.-Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics* **2010**, 12, 12899-12908.
- [54].Wang, R.; Lu, Y.; Wang, S., Comparative evaluation of 11 scoring functions for molecular docking. *Journal of medicinal chemistry* **2003**, 46, 2287-2303.
- [55].Li, Y.; Yang, J., Structural and sequence similarity makes a significant impact on machine-learning-based scoring functions for protein–ligand interactions. *Journal of chemical information and modeling* **2017**, 57, 1007-1012.
- [56].Gabel, J.; Desaphy, J.; Rognan, D., Beware of Machine Learning-Based Scoring Functions • On the Danger of Developing Black Boxes. *Journal of chemical information and modeling* **2014**, 54, 2807-2815.
- [57].Shen, C.; Ding, J.; Wang, Z.; Cao, D.; Ding, X.; Hou, T., From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, 10, e1429.
- [58].Foloppe, N.; Hubbard, R., Towards predictive ligand design with free-energy based computational methods? *Current medicinal chemistry* **2006**, 13, 3583-3608.

- [59].Guedes, I. A.; Pereira, F. S. S.; Dardenne, L. E., Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Frontiers in pharmacology* **2018**, 9, 1089.
- [60].Crespo, A.; Rodriguez-Granillo, A.; Lim, V. T., Quantum-mechanics methodologies in drug discovery: applications of docking and scoring in lead optimization. *Current topics in medicinal chemistry* **2017**, 17, 2663-2680.
- [61].Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie International Edition* **2003**, 42, 2718-2736.
- [62].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. *Journal of computational chemistry* **2009**, 30, 2785-2791.
- [63].Salo-Ahen, O. M. H.; Alanko, I.; Bhadane, R.; Bonvin, A. M. J. J.; Honorato, R. V.; Hossain, S.; Juffer, A. H.; Kabedev, A.; Lahtela-Kakkonen, M.; Larsen, A. S., Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. *Processes* **2021**, 9, 71.
- [64].Leimkuhler, B.; Matthews, C., *Molecular Dynamics*. Springer: 2016.
- [65].Rahman, A., Correlations in the motion of atoms in liquid argon. *Physical review* **1964**, 136, A405.
- [66].Verlet, L., Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **1967**, 159, 98.
- [67].The nobel Prize. The Nobel Prize in Chemistry 2013. <https://www.nobelprize.org/prizes/chemistry/2013/advanced-information/> (June 22th, 2021),
- [68].Adcock, S. A.; McCammon, J. A., Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews* **2006**, 106, 1589-1615.
- [69].Wang, D.-B.; Hsiao, F.-B.; Chuang, C.-H.; Lee, Y.-C., Algorithm optimization in molecular dynamics simulation. *Computer Physics Communications* **2007**, 177, 551-559.
- [70].Sagui, C.; Darden, T. A., Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **1999**, 28, 155-179.

- [71].Koradi, R.; Billeter, M.; Güntert, P., Point-centered domain decomposition for parallel molecular dynamics simulation. *Computer Physics Communications* **2000**, 124, 139-147.
- [72].Alfe, D., Ab initio molecular dynamics, a simple algorithm for charge extrapolation. *Computer Physics Communications* **1999**, 118, 31-33.
- [73].González, M. A., Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **2011**, 12, 169-200.
- [74].Guvench, O.; MacKerell, A. D., Comparison of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins* **2008**, 63-88.
- [75].Monticelli, L.; Tieleman, D. P., Force fields for classical molecular dynamics. *Biomolecular simulations* **2013**, 197-213.
- [76].Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Advances in protein chemistry* **2003**, 66, 27-85.
- [77].Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K., Comparing molecular dynamics force fields in the essential subspace. *PLoS One* **2015**, 10, e0121114.
- [78].Ricci, C. G.; de Andrade, A. S. C.; Mottin, M.; Netz, P. A., Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions. *The Journal of Physical Chemistry B* **2010**, 114, 9882-9893.
- [79].Platania, C. B. M.; Bucolo, C. Molecular Dynamics Simulation Techniques as Tools in Drug Discovery and Pharmacology: A Focus on Allosteric Drugs. In *Allostery*; Springer: 2021, pp 245-254.
- [80].Paquet, E.; Viktor, H. L., Computational methods for Ab initio molecular dynamics. *Advances in Chemistry* **2018**, 2018, 9839641.
- [81].Hollingsworth, S. A.; Dror, R. O., Molecular dynamics simulation for all. *Neuron* **2018**, 99, 1129-1143.
- [82].Mouchlis, V. D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A. G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G., Advances in de novo drug design: From conventional to machine learning methods. *International journal of molecular sciences* **2021**, 22, 1676.
- [83].Schneider, G.; Baringhaus, K. H., De novo design: from models to molecules. *De novo Molecular Design* **2013**, 1-55.
- [84].Schneider, G.; Fechner, U., Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery* **2005**, 4, 649-663.

- [85].Kawai, K.; Nagata, N.; Takahashi, Y., De novo design of drug-like molecules by a fragment-based molecular evolutionary approach. *Journal of chemical information and modeling* **2014**, 54, 49-56.
- [86].Liu, X.; Ijzerman, A. P.; van Westen, G. J. P. Computational Approaches for De Novo Drug Design: Past, Present, and Future. In *Artificial Neural Networks*; Springer: 2021, pp 139-165.
- [87].Abraham, A.; Nedjah, N.; de Macedo Mourelle, L. Evolutionary computation: from genetic algorithms to genetic programming. In *Genetic systems programming*; Springer: 2006, pp 1-20.
- [88].Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M. A.; Steinbeck, C., COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics* **2021**, 13, 1-13.
- [89].Meyers, J.; Fabian, B.; Brown, N., De novo molecular design and generative models. *Drug Discovery Today* **2021**.
- [90].Krishnan, S. R.; Bung, N.; Bulusu, G.; Roy, A., Accelerating de novo drug design against novel proteins using deep learning. *Journal of Chemical Information and Modeling* **2021**, 61, 621-630.
- [91].Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R., Interpretable Deep Learning for De Novo Design of Cell-Penetrating Abiotic Polymers. *bioRxiv* **2020**.
- [92].Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T., The rise of deep learning in drug discovery. *Drug discovery today* **2018**, 23, 1241-1250.
- [93].Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A., Reinforced adversarial neural computer for de novo molecular design. *Journal of chemical information and modeling* **2018**, 58, 1194-1204.
- [94].Born, J.; Manica, M.; Oskoei, A.; Cadow, J.; Markert, G.; Martínez, M. R., PaccMannRL: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. *Isience* **2021**, 24, 102269.
- [95].Ståhl, N.; Falkman, G.; Karlsson, A.; Mathiason, G.; Bostrom, J., Deep reinforcement learning for multiparameter optimization in de novo drug design. *Journal of chemical information and modeling* **2019**, 59, 3166-3176.

- [96].Popova, M.; Isayev, O.; Tropsha, A., Deep reinforcement learning for de novo drug design. *Science advances* **2018**, 4, eaap7885.
- [97].Hartenfeller, M.; Schneider, G., De novo drug design. *Chemoinformatics and computational chemical biology* **2010**, 299-323.

The background of the page is a light gray abstract pattern. It features a complex network of thin lines connecting small dots, resembling a molecular structure or a data network. Interspersed within this network are various geometric shapes, primarily hexagons, some of which are solid gray and others are outlined. The overall effect is a technical, scientific, or digital aesthetic.

CAPÍTULO II

O século XXI tem sido chamado de era da big data. Até o ano de 2020, o volume de dados atingiu 59 zettabytes (ZB), esperando-se alcançar 149 ZB em 2024 (um ZB equivale a 1×10^{12} gigabytes) [1]. Na área da química medicinal, essa situação não é muito diferente; a informação sobre produtos naturais e seus metabólitos secundários, como ponto de partida para o desenvolvimento de novos tratamentos para inúmeras doenças, aumenta exponencialmente, sendo fundamental o desenvolvimento de bancos de dados - entre eles ZINC, PubChem e ChEMBL, que possuem milhões de estruturas - para o correto armazenamento dessas informações [2].

A Asteraceae é uma das famílias botânicas com maior diversidade no mundo, possuindo mais de 32.000 espécies identificadas pertencentes a 1.900 gêneros [3, 4]. Várias espécies desta família demonstraram forte atividade antioxidante, anti-inflamatória e antimicrobiana, além de propriedades diuréticas e cicatrizantes. Seus efeitos farmacológicos podem ser atribuídos a uma variedade de compostos fitoquímicos, incluindo polifenóis, ácidos fenólicos, flavonoides, acetilenos e triterpenos [5].

Com tanta variedade e quantidade de informações, a classificação sistemática destas por meio de bancos de dados torna-se cada vez mais relevante para estudos futuros com a Asteraceae, e ferramentas específicas como SistemX e AsterDB, desenvolvidas no Brasil, tornam-se essenciais para o desenvolvimento de estudos quimioinformáticos, favorecendo a obtenção de informações relevantes e não redundantes [2].

As famílias de plantas mais promissoras para identificar compostos com potencial atividade leishmanicida são Asteraceae e Lamiaceae [6]. No entanto, várias classes de metabólitos secundários, assim como alvos terapêuticos, ainda não foram estudadas em profundidade. Para direcionar os estudos computacionais observados nos próximos capítulos do presente trabalho, fizemos esta revisão de literatura, procurando a maior parte dos estudos que utilizaram ferramentas computacionais para examinar vários compostos identificados na família Asteraceae na busca por possíveis candidatos a medicamentos contra *Leishmania*, enfatizando as possíveis bases de dados que podem ser utilizadas, assim como as principais espécies e alvos terapêuticos usados.

. REFERÊNCIAS

- [1]. SULOVA, Snezhana et al. big data processing in the logistics industry. *Economics and computer science*, n. 1, p. 6-19, 2021.
- [2]. SCOTTI, Marcus Tullius et al. SistematiX, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules*, v. 23, n. 1, p. 103, 2018.
- [3]. BOHM, Bruce A.; STUESSY, Tod F. **Flavonoids of the sunflower family (Asteraceae)**. Springer Science & Business Media, 2001.
- [4]. MOHANTA, Yugal Kishore et al. Potential use of the Asteraceae family as a cure for diabetes: A review of ethnopharmacology to modern day drug and nutraceuticals developments. **Frontiers in Pharmacology**, v. 14, 2023.
- [5]. ROLNIK, Agata; OLAS, Beata. The plants of the Asteraceae family as agents in the protection of human health. **International journal of molecular sciences**, v. 22, n. 6, p. 3009, 2021.
- [6]. CORTES, Sofia et al. Potential of the natural products against leishmaniasis in Old World-a review of in-vitro studies. **Pathogens and global health**, v. 114, n. 4, p. 170-182, 2020.

Discovery of Alternative Chemotherapy Options for Leishmaniasis via Computational Studies of Asteraceae

Chonny Herrera-Acevedo¹, Camilo Perdomo-Madrigal², Eugene N. Muratov¹, Luciana Scotti¹, Marcus Tullius Scotti^{1,*}.

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil.

² School of Science, Universidad de Ciencias Aplicadas y Ambientales, Calle 222 # 55–37, Bogotá D.C., Colombia.

*Contact author. Email: mtscotti@gmail.com

Artigo publicado em ChemMedChem (IF: 3.540)

DOI Number: <https://doi.org/10.1002/cmdc.202000862>

Ano de publicação: 2021

Abstract: Leishmaniasis is a complex disease caused by over 20 *Leishmania* species that primarily affect populations with poor socioeconomic conditions. Currently available drugs for leishmaniasis treatment include amphotericin B, paromomycin, and pentavalent antimonials, which have been associated with several limitations, such as low efficacy, the development of drug resistance, and high toxicity. Natural products are an interesting source of new drug candidates. The Asteraceae family includes more than 23,000 species worldwide. Secondary metabolites that can be found in species from this family have been widely explored as potential new treatments for leishmaniasis. Recently, computational tools have become more popular in medicinal chemistry to establish experimental designs, identify new drugs, and compare the molecular structures and activities of novel compounds. Here, we review various studies that have used computational tools to examine various compounds identified in the Asteraceae family in the search for potential drug candidates against *Leishmania*.

Keywords: Asteraceae · Leishmaniasis · *In Silico* · Natural Products · Databases.

1. Introduction

Leishmaniasis is a general term used to describe more than 20 diseases caused by protozoans of the genus *Leishmania* (31 species are known to be mammalian parasites, of which 20 species are pathogenic in humans), which are transmitted by various species of phlebotomine sandflies (Diptera and Psychodidae) [1,2]. These diseases are categorized as neglected tropical diseases (NTDs) and represent a huge challenge for developing countries due to their economic and social impacts. The risk of infection and disease severity are associated with the socioeconomic conditions of the population, including the sanitation and nutritional conditions of a community, environmental or climate changes, and the response of the host immune system to the infection [2]. Leishmaniasis can be classified according to the severity of the infection as visceral (VL), cutaneous (CL), and mucocutaneous (MCL) leishmaniasis [3,4].

Similar to other NTDs, leishmaniasis does not attract much attention from most pharmaceutical companies because the research and production of new medicines to treat these diseases are unlikely to result in significant profit [5]. This situation has prompted the cooperation between governments, global organizations, and research institutions to address these diseases, which represent public health challenges in several countries [6,7].

In attempts to identify low-cost medicines for the treatment of leishmaniasis, many researchers have focused their investigations on natural products, which, in contrast to conventional chemotherapeutic drugs, can offer high efficacy and low toxicity [8]. Currently, more than 50% of all new drugs approved by the United States Food and Drug Administration (US FDA) are associated with natural products [9]. In recent decades, extracts and fractions of plants in the Asteraceae, Lamiaceae, Apiaceae, and other families have been used in traditional medicine approaches to treat leishmaniasis, showing high activity and selectivity [10]. Asteraceae is one of the most well-studied plant families for the treatment of parasite-associated diseases [8]. For instance, artemisinin, a sesquiterpene lactone (SL) derived from *Artemisia annua*, has been approved for the treatment of malaria due to its endoperoxidase properties [11,12].

The high variability of *Leishmania* strains and their strong resistance to currently available drugs are currently the biggest obstacles to the search for new antileishmanial treatments. Since the 1950s, anti-*Leishmania* treatments have primarily been based on pentavalent antimonial compounds (Sb^{V}), including meglumine antimoniate (Glucantime®) and sodium stibogluconate (Pentostam®), although the efficacy of these two Sb^{V} treatments has declined over six decades of use, and multiple studies have reported

several complications associated with these two drugs, such as high toxicity, parasite resistance, prolonged treatments, and multiple side effects [8,13,14]. More recent chemotherapies that have been developed against different types of *Leishmania*, including amphotericin B, miltefosine, and paromomycin, have been reported to present similar problems [15,16]. Therefore, understanding the mechanism of action through which available drugs exert leishmanicidal effects and the mechanism through which parasites develop antibiotic resistance is extremely important. The discovery of new antibiotic targets would facilitate the identification of more effective and selective leishmanicidal drugs [8,17].

Computer-aided drug design (CADD) offers the possibility to design new drugs *in silico*, which, combined with conventional wet-lab strategies, allows researchers to develop effective treatments in shorter times and at lower costs. Specifically, quantitative structure-activity relationship (QSAR) models can be used to identify molecular patterns that can be modified in a molecular motif to maximize activity [18]. Molecular docking methodology can also be used to explore the behaviors of small molecules at the binding site of a target protein [19].

This review summarizes existing databases (DBs) associated with Asteraceae and explores the reported results of studies that have used *in silico* methodologies (particularly machine learning and molecular docking calculations) to identify new structures with potential anti-*Leishmania* activities, based on secondary metabolites found in Asteraceae species.

2. Databases as key tools for cheminformatic studies exploring Asteraceae components.

Asteraceae is one of the more-studied plant families in the world. Asteraceae are highly distributed worldwide, including 1,000 genera and 25,000–30,000 species [20]. More than 22 million georeferenced reports have been identified (Figure 1)) [21]. Additionally, secondary metabolites derived from members of this family have successfully demonstrated antiparasitic activity, such as artemisinin (*Artemisia annua*), an SL with antimalarial activity that received the Nobel Prize in Medicine in 2015 and has been approved for treatment against malaria caused by *Plasmodium falciparum* [11,12].

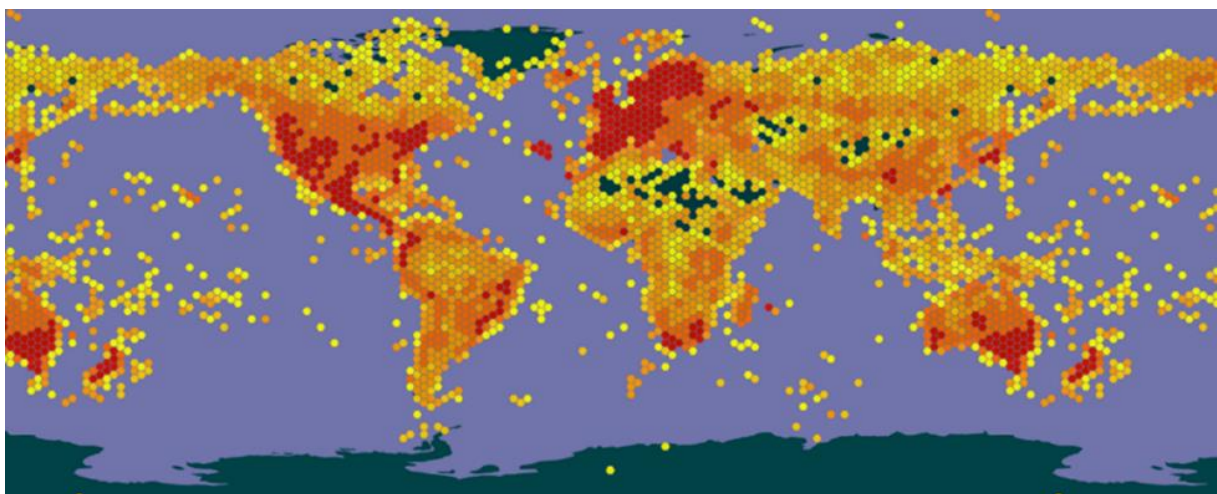


Figure 1. Asteraceae is highly distributed worldwide. More than 23,000 species and 22 million georeferenced reports are registered in the Global Biodiversity Information Facility (GBIF) [21].

Due to the large amounts of existing data for Asteraceae, the systematic storage and retrieval of information regarding identified metabolites in this family are important. Natural products DBs [22,23], allow data to be classified according to chemical structures, trade names, systematic names, synonyms, molecular formulae, and a wide range of calculated and/or experimental data [24].

Currently, natural product DBs are key tools for CADD studies, including Dictionary of Natural Products [25], NUBBEdb [26], Super Natural II [27], NAPRALERT [28], KNApSACK Family Databases [29], among others.

DNP is one of the most complete natural product DBs currently available, providing chemical, physical, and structural data for over 190,000 natural products organized into more than 43,000 entries; however, DNP is a commercial DB and is relatively expensive to access, which limits its use as a research tool [30,31]. KNApSACK Metabolomics contains more than 10,000 structures, allowing metabolites to be searched according to mass spectrometry (MS) peaks, molecular weights, molecular formulas, and species [29].

SuperNatural II, a free, web-based natural product DB, contains approximately 326,000 molecules. NUBBEdb, another web-based DB, has registered a variety of secondary metabolite classes from among the biodiversity in Brazil, including botanical, chemical, pharmacological, and toxicological compound data [26]. NAPRALERT was developed by the University of Illinois at Chicago and is described as a relational DB of natural products, including ethnomedical and pharmacological/biochemical data for

extracts tested in a variety of organisms using *in vitro*, *in situ*, and *in vivo* studies, in addition to studies performed in humans, such as case reports, non-clinical trials, and clinical studies [28].

None of the above-mentioned DBs are specifically focused on the Asteraceae family. In that past decade, two natural products DBs have been introduced, containing data for a large number of secondary metabolites from Asteraceae: Sistematrix and AsterDB. Studies using the compounds registered in these two DBs have recently been published, including a combined virtual screening approach aiming to identify antichagasic structures [23] and a fingerprinting metabolomics study of tropical mistletoes that grow on aluminum-accumulating and aluminum-excluding hosts [32].

Sistematrix (<http://sistematrix.ufpb.br>) was developed in the cheminformatics laboratory of the Federal University of Paraíba, whose interface includes the following aspects: (a) the ability to search by structure, SMILES (Simplified Molecular-Input Line-Entry System) code, compound name, and species; (b) the ability to save chemical structures identified during searches; (c) compound data results, including important natural products chemical characteristics, such as spectrometric data; and (d) the ability to identify specific information regarding the taxonomic rank (from family to species) of the plant from which each compound was isolated, the searched-for molecule, spectroscopy data, bibliographic references, and Global Positioning System (GPS) coordinates. Sistematrix includes more than 1,300 SLs and 850 flavonoids and chalcones that are associated with more than 4,000 botanical occurrences in Asteraceae [22].

AsterDB (<http://www.asterbiochem.org/asterdb>) is the AsterBioChem *in-house* DB, which contains hundreds of chemical structures that have only been reported in Asteraceae species. AsterDB is the first DB to provide free access and is dedicated specifically to this botanical family. At this time, AsterDB has registered 2,500 unique chemical structures of terpenoids, flavonoids, trans-cinnamic acid derivatives, and other minor chemical classes of natural products, including more than 1,000 structures of SLs [33].

3. Computational studies against *Leishmania*.

For *Leishmania*, compounds from diverse plant sources and their modified analogs have been tagged with nanoparticles to explore potential mechanisms to increase the delivery, efficacy, and bioavailability of these compounds [34]. The leishmanicidal activity of Asteraceae has been widely studied, including multiple ethanol extracts [35], SLs (which are characteristic Asteraceae chemomarkers) [8], flavonoids [36,37], tannins and

steroids. Some secondary metabolites found to efficiently treat experimental leishmaniasis include ferulic acid, rosmarinic acid, and ursolic acid [38].

In this section, we summarized key computational studies that have been based on compounds identified in Asteraceae species and used to identify new structures with potential activity against various *Leishmania* species (Figure 2). These results are divided into two groups: those that used machine learning methodologies and those that were based on the structure of a target protein from *Leishmania* (Table 1), primarily through the application of molecular docking calculations.

3.1. Machine learning studies.

Hologram QSAR (HQSAR) is a technique that employs specialized fragment fingerprints (called molecular holograms) as variables for the prediction of biological activities or other structurally related data [39]. Trossini et al. built HQSAR models using 16 series of fragment distinctions and fixed fragment sizes (4–7 atoms) based on 40 SLs that were identified in several species of Asteraceae and showed antiprotozoal activity against four NTDs, including *Leishmania donovani*, and cytotoxicity against L6 rat skeletal myoblasts. Cross-validation (leave-one-out and leave-n-out) and external validation, using different latent variables, were used to test the biological activity prediction capabilities of these models [40].

The best HQSAR models for activity against *L. donovani* ($Q^2 = 0.775$) and cytotoxicity ($Q^2 = 0.647$) showed that SL activities against *L. donovani* and L6 cytotoxicity are clearly influenced by stereoselectivity and H-bond interactions because these models were constructed using only these two parameters. The authors also observed that the oxygen atom of the oxirane group contributes negatively to the *L. donovani* HQSAR model and that the α , β -unsaturated groups are fundamental to the biological activity of SLs [40,41].

The antileishmanial activity of seventeen SLs, isolated from five species of the tribe Vernonieae, were tested *in vitro* using the parasitic promastigote forms of *Leishmania braziliensis* and *Leishmania amazonensis*. Isodeoxyelephantopin (1, half-maximal inhibitory concentration (IC₅₀): 1.45 μ M) and deoxyelephantopin (2, IC₅₀: 1.34 μ M), two germacranolides that were previously isolated from *Elephantopus carolinianus* [42] were identified as the most active structures against *L. braziliensis*, whereas centratherin, with an IC₅₀ value of 1.45 μ M, had the highest activity against *L. amazonensis* [43] (Figure 2).

To establish the structural features associated with the observed IC₅₀ values, various molecular descriptors were calculated using the PaDel-Descriptor software [44]. The constructed QSAR models showed that descriptors related to the partition coefficient (LogP) and polarizability (bpol) were the best-correlated descriptors with determination coefficients (R^2), with R^2 values greater than 0.77 for *L. braziliensis* and greater than 0.81 for *L. amazonensis*. Validation data for the QSAR models were not shown. The analysis of these QSAR models showed an optimal range for LogP, at the point where the hydrophilic and lipophilic properties are balanced, which was correlated with an increase in the biological activity. The three most active molecules had the lowest polarizability, based on bpol values, which is associated with the ability of these molecules to penetrate the cellular membrane and may affect the van der Waals interactions involved in ligand–receptor recognition.

In a study reported by Tasdemir et al., the antileishmanial activity of a series of 105 flavonoid aglycones and glycosides was tested against *L. donovani* axenic amastigotes. The majority of the tested metabolites presented considerable leishmanicidal potential, with fisetin (compound 3), luteolin (compound 4), 3-hydroxyflavone (compound 5), and quercetin (compound 6) representing the most potent anti-*L. donovani* structures, with IC₅₀ values of 2.1, 2.8, 2.9, and 3.3 μ M, respectively [45]. Compounds 4, 5, and 6 had only slight or no toxicity in mammalian cells, with selectivity index (SI) values from 20 to 64. These three molecules have been reported as common flavonoids found in the Asteraceae family, which display antiparasitic activity [46].

Using Molecular Operating Environment (MOE) software [47], 241 molecular descriptors were calculated using the lowest-energy conformer of each compound, and a partial-least squares (PLS) analysis was performed. A low correlation value for *L. donovani* was obtained. The authors proposed three possibilities to explain these results, including that a common structure-activity relationship (SAR) may not exist for the studied compounds, that the range of biological data (the difference between the most and least active was relatively small for *L. donovani*) may have been too small, and that the structural factors underlying these biological effects may not be represented by the chosen molecular descriptions. Despite these contradictory results, this study presented a good computational approximation for establishing a relationship between the *in vitro* and *in silico* studies of some flavonoids (although most studies are performed using SLs) found in the Asteraceae family and served as a starting point for the rational design of quercetin (compound 6) derivatives as potent leishmanicidal agents [45].

Chibli et al. performed two studies involving *Leishmania major* dihydroorotate dehydrogenase (*LmDHODH*) and *in silico* methods to explore natural products from Asteraceae [48,49]. *LmDHODH* is a flavoenzyme that catalyzes the stereoselective oxidation of (S)-dihydroorotate to orotate in the fourth of the six conserved, enzymatic reactions involved in the *de novo* pyrimidine biosynthetic pathway, which is a protein involved in vital cellular functions [50].

The *in vitro* inhibition of *LmDHODH*, including cross-validation against human DHODH (*HsDHODH*), was examined for 57 natural products, including 48 that have previously been reported in Asteraceae species. IC_{50} values ranging from 27 to 1,200 μ M were obtained against *LmDHODH*, and the inhibition was found to be highly selective, as no relevant inhibitory effect was observed against *HsDHODH* [49].

A QSAR model based on molecular descriptors and a pharmacophore-based 3D-QSAR were performed in MOE software [47]. Variable selection, with a genetic algorithm (GA) and multiple linear regression (MLR), followed by partial least squares (PLS) regression were performed, which were validated by calculating R^2 , Q^2_{CV} , and Q^2_{EXT} . In the first QSAR model, a set of 21 SLs were used to build the model, and both 2-dimensional (2D, $n = 198$) and 3-dimensional (3D, $n = 126$) molecular descriptors were calculated for each molecule. The 21 SLs were divided into a training set (15 SLs) and a test set (6 SLs). The results showed validation parameters for R^2 of 0.83, Q^2_{CV} of 0.69, and Q^2_{EXT} of 0.66 for the best linear model obtained. The descriptors that influence the model are associated with the presentation of hydrophobic regions across the molecular surface, in addition to the increased width and lower hydrophobicity of the molecules (Only 3D descriptors, *vsurf_ID8* explain 58% of the variance in activity values) [49].

Two of the most-active SLs (both present in Asteraceae), 2-oxo-8b-tigloyloxyguaia-1(10),3,11(13)-trien-6a,12-olide (compound 7, $IC_{50} = 27$ mM) and glaucolide B (compound 8, $IC_{50} = 31$ mM) were examined in a pharmacophore-based QSAR model. Using this method, the SLs were aligned based on the most important pharmacophore descriptors, including cyclopentenone, tigloyl, the carbonyl oxygen of cyclopentenone, and the ester carbonyl oxygen of tigloyl moieties for compound 7, and the acetyl group at CH₂-13, acetyl, the ester carbonyl oxygen of the acetyl group at CH₂-13, and the acetyl groups for compound 8. The best 3D-QSAR models obtained using the pharmacophore descriptors (R^2 : 0.72; Q^2_{CV} : 0.50 and Q^2_{EXT} : 0.62) confirmed the importance of the correct ligand orientation and the molecular surface features to induce stronger inhibition, which suggested shared properties for a putative common binding site [49].

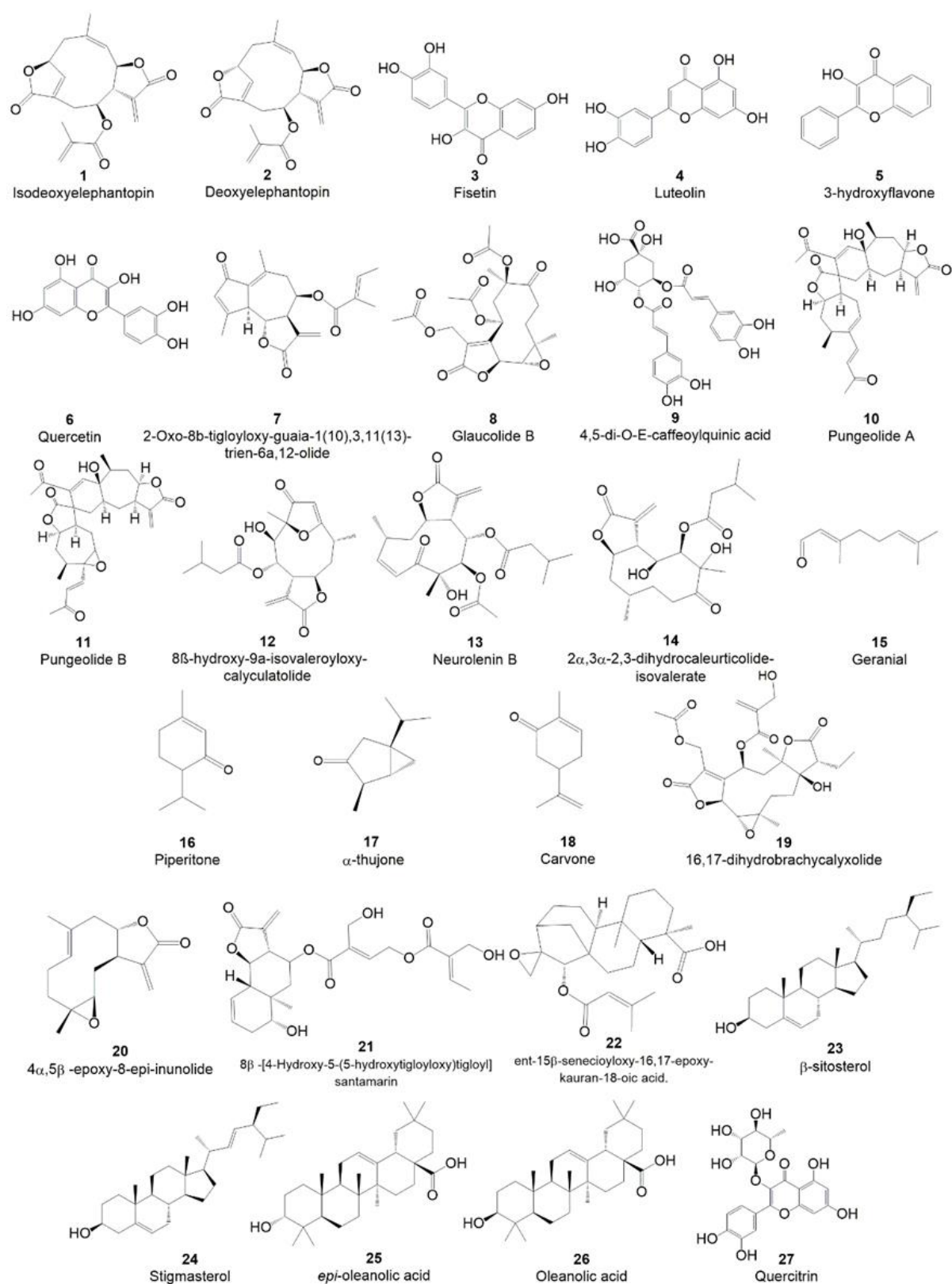


Figure 2 Asteraceae secondary metabolites were identified as *hits* based on *in silico* methodologies.

In a metabolomic study, IC_{50} values against *LmDHODH* were determined for extracts from 59 Asteraceae species, which ranged from 0.148 mg/mL to 9.4 mg/mL. Using SIMCA-P+ Software (v.13.0, Umetrics, Umea, Sweden), unsupervised principal component analysis (PCA) and supervised orthogonal projections to latent structures discriminant analysis (OPLS-DA) were performed to discover structures with inhibitory capabilities against *LmDHODH*. Classification labels of active (IC_{50} less than 500 μ g/mL) and inactive (IC_{50} greater than 500 μ g/mL) were assigned to separate the extracts, which resulted in a clear differentiation into two clusters, indicating significant differences in the chemical compositions of these two groups. The dereplication of the metabolic fingerprints for the active extracts resulted in the identification of 48 metabolites, including one belonging to the quinic acids class: 4,5-di-O-E-caffeoylquinic acid (compound 9, IC_{50} 73 μ M), which demonstrated inhibitory capacity against *LmDHODH* [48].

A second OPLS-DA model was built based on the previously described model to predict the *LmDHODH* inhibitory potential of extracts that were not tested in this study. The model was validated with a training set (44 extracts) and a test set (15 extracts). The results showed validation parameters of R^2 : 0.96, Q^2 : 0.72, and $RMSE_{CV}$: 0.24, with a 60% correct prediction rate for the test set. The good predictability of this model suggests that it may represent a useful tool for identifying active extracts against this protein in *Leishmania* and may serve as a good starting point for the development of leishmanicidal treatments [48].

Recently, promising enzyme-targeting, antileishmanial SLs from Sistemax that may act against *L. donovani* were identified in a combined approach, based on two different virtual screening (VS) methods. A ChEMBL dataset, including 3,159 and 1,569 structures that were previously tested against *L. donovani* amastigotes and promastigotes *in vitro*, respectively, was used to develop two random forest models, which performed with greater than 74% accuracy in both the cross-validation and test sets [51].

A ligand-based VS assay was performed against the 1,306 Sistemax-registered SLs. In parallel, using the crystal structures of three *L. donovani* target proteins, N-myristoyltransferase (PDB ID: 2WUU) [52], ornithine decarboxylase (PDB ID: 2OOO) [53], and mitogen-activated protein kinase 3 (PDB ID: 4O2Z), and a homology model of pteridine reductase 1 were used to perform a structure-based VS (molecular docking) of the entire Sistemax SL dataset. The consensus analysis of these two VS approaches resulted in the normalization of probability scores and identified 13 promising, enzyme-targeting, antileishmanial SLs from Sistemax that may act against *L. donovani* [51]. Despite this,

the work is exploratory, since *in vitro* tests showing leishmanicidal activity of these molecules were not performed. Further studies validating experimentally the obtained results are necessary.

3.2. Molecular docking studies.

Bernal and Coy-Barrera performed a molecular docking study examining 123 sesquiterpene-related compounds that possess *in vitro* antiparasitic activity within the active sites of four targets (Table 1): *L. major* pteridine reductase-1 [PTR1, Protein Data Bank (PDB) IDL 2QHX] [54], cysteine synthase (CS, PDB ID: 4AIR) [55], trypanothione synthetase (TS, PDB ID: 2VOB) [56] and *L. donovani* (NMT, PDB: ID 2WUU) [52]. All calculations were performed using AutoDock/Vina (1.1.2). After, PCA analysis, the drug-like properties were calculated from the best-docked sesquiterpenes, using R software and ChemBio3D (Cambridge Soft Corporation, Cambridge, MA, USA) [57]. For all ligands, molecular energy minimization was performed using Merck Molecular Force Field (MMFF94) included in the Spartan '14 software with a limit of 500 conformers. Energetically lowest stable conformers within a 6 kcal/mol energy range were geometrically optimized using the semi-empirical AM1 parametrization.

The docking results showed that dimeric xanthanolide compounds, including pungiolide A (compound **10**) and pungiolide B (compound **11**), exhibited the best affinity values (−10.6 kcal/mol) for *L. major* PTR1, with even lower values than that for the redocked ligand (DB07765, −10.2 kcal/mol). Compound **10** was also one of the best-docked structures within *L. major* CS, with an affinity value of −10.6 kcal/mol, which indicated a higher affinity for this target than the inhibitor used as a control (ZINC01690699, −9.6 kcal/mol). Two germacranolides, 8 β -hydroxy-9 α -isovaleroyloxy-calyculatolide (compound **12**, −8.7 kcal/mol) and neurolenin B (compound **13**, −8.5 kcal/mol) were the best-docked secondary metabolites from Asteraceae for *L. major* TS, with similar values as those identified for the control ligand (DDD66604, −8.2 kcal/mol).

L. donovani NMT1 was the most restrictive protein of the four tested, with none of the 123 sesquiterpenoids displaying higher affinity than the control ligand (DDD64558; −8.0 kcal/mol). However, 2 α ,3 α -2,3-dihydro caleurticolide-isovalerate (compound **14**), together with compounds **10** and **11**, were three of the ten-best docked sesquiterpenoids, with affinity values close to −6.4 kcal/mol [57]. Compounds **10** and **11** interacted within the active site of *L. major* PTR1 at the same residues, establishing two H-bonds with the side chain of R17 and two H-bonds with the amino acid backbone of S111 and S227. The

α,β -unsaturated groups, such as 2-methylen- γ -lactone, were able to interact with the residues of the pocket, satisfying the structural requirements for a possible Michael addition. Through PCA analysis, the vina scores for *L. major* PTR1 and CS and *L. donovani* NMT were found to be highly correlated, and compounds **10** and **11** were differentiated compared with the other sesquiterpenoids, demonstrating poor affinity for *L. major* TS but good results for the other three targets. Finally, compound **14** presented one of the best lead-like properties. These screening results indicated that this method could be used to identify ligands with potential activity against *Leishmania*. However, only structures with promising docking scores were identified, *in vitro* tests were not performed [57].

Ogungbe and Setzer performed a study with the aim of identifying potential *Leishmania* biochemical targets among known plant-derived isoprenoids that have exhibited *in vitro* antiparasitic activity. For this purpose, molecular docking calculations were performed in Molegro Virtual docker (v.5.0. Molegro ApS, Aarhus, Denmark), using 29 PDB crystal structures for targets from different species of *Leishmania*: *L. major* (12 proteins), *L. mexicana* (8 proteins), *L. donovani* (4 proteins), and *L. infantum* (5 proteins) Each ligand structure was built using Spartan '10 for Windows and the structures were geometry optimized using the MMFF force field [58]. Within the group of tested natural products, which can be found in the Asteraceae family, monoterpenoids were found to selectively dock with *L. infantum* nicotinamidase (PDB ID: 3R2J) [59], *L. major* uridine diphosphate-glucose pyrophosphorylase (PDB ID: 2OEF and 2OEG) [60] and methionyl t-RNA synthetase (PDB ID: 3KFL) [61]; germacranolides sesquiterpenoids docked with *L. major* methionyl t-RNA synthetase (PDB ID: 3KFL) [61] and dihydroorotate dehydrogenase (PDB ID: 3MHU) [62]; diterpenoids docked with *L. mexicana* glycerol-3-phosphate dehydrogenase (PDB ID: 1N1E) [63], and triterpenoids docked with *L. infantum* sterol 14 α -demethylase (CYP51, PDB ID: 3L4D) [64].

Geranial (compound **15**), a monoterpene found in the essential oils derived from several species of Asteraceae [65,66] exhibited the lowest docking energies for *L. major* uridine diphosphate-glucose pyrophosphorylase (−76.9 kJ/mol) and *L. major* methionyl t-RNA synthetase (−76.8 kJ/mol). Piperitone (compound **16**), found in the essential oils of several common species of Asteraceae, such as White wormwood (*Artemisia herba alba*) [67] and Chamomile (*Matricaria chamomilla*) [68] also showed a high affinity for *L. major* uridine diphosphate-glucose pyrophosphorylase (−68.0 kcal/mol) and *L. infantum* nicotinamidase (−73.0 kJ/mol). Together with α -thujone (compound **17**, −74.5 kJ/mol) and

carvone (compound **18**, -73.6 kJ/mol), these molecules presented docking energies even less than that of the co-crystallized ligand, nicotinic acid, which had a docking energy of -65.5 kcal/mol [58].

Among the sesquiterpenes, 16,17-dihydrobrachycalixolide (compound **19**), was the strongest-docking germacranolide, with the lowest energies against *L. major* methionyl t-RNA synthetase (-152.9 kJ/mol) and dihydroorotate dehydrogenase (-129.1 kJ/mol). This energy docking was lower than that for the PDB ligand, 5-nitroorotic acid (-102.2 kJ/mol). Additionally, 4 α ,5 β -epoxy-8-epi-inunolide (compound **20**), which is found in multiple Asteraceae species, such as in the aerial segments of *Stevia ovata* [69], showed docking selectivity toward both *L. major* methionyl t-RNA synthetase (-99.1 kJ/mol) and dihydroorotate dehydrogenase (-100.5 kJ/mol). 8 β -[4-Hydroxy-5-(5-hydroxytigloyloxy)tigloyl]santamarin (compound **21**) was the strongest-docking eudesmanolide, presenting docking selectivity for *L. major* methionyl t-RNA synthetase (-153.1 kJ/mol) [58].

The most selective targets for diterpenoids and triterpenoids were *L. mexicana* glycerol-3-phosphate dehydrogenase and *L. infantum* sterol 14 α -demethylase, respectively. Kaurane diterpenoids were docked to this target, especially *ent*-15 β -seneciyoxy-16,17-epoxy-kauran-18-oic acid (compound **22**), which was the best-docked kaurane (-128.9 kJ/mol). Compound **22** can be found in Asteraceae species and was initially isolated during a bioactivity-guided fractionation of the total crude extract of *Aspilia pruliseta* [70]. Finally, triterpenes and steroids demonstrated significant docking preferences for *L. infantum* CYP51, with a range of energy values from (-120 to -110 kJ/mol), which highlighted molecules that are reported in Asteraceae plants such as β -sitosterol (compound **23**) and stigmasterol (compound **24**) [71].

So, from the fact that this study uses molecules that have previously been shown to have *in vitro* activity, this research is an advance looking for alternative chemotherapies against *Leishmania* parasites, establishing a possible mechanism of action for these secondary metabolites obtained from species of Asteraceae

Table 1. Enzymes from different *Leishmania* species were studied by molecular docking, using secondary metabolites from Asteraceae.

Species	Enzyme	PDB ID	Resolution [Å]	Reference
<i>L. donovani</i>	Cathepsin B	Homologue model	-	Ogungbe et al. [58]
	Cyclophilin A	2HAQ [72]	1.97	
		3EOV [73]	2.60	
	Dihydroorotate dehydrogenase	3C61	1.80	
	Mitogen-activated protein kinase 3 - MPK3	4O2Z	2.71	Herrera-Acevedo et al. [51]
	Ornithine decarboxylase	2O00 [53]	1.90	
	Pteridine reductase 1	Homologue model	-	
	N-Myristoyl transferase	2WUU [52]	1.42	Bernal, Coy-Barrera [57]
				Ogungbe et al. [58]
				Herrera-Acevedo et al. [51]
<i>L. infantum</i>	14-alpha demethylase (CYP51)	3L4D [64]	2.75	Warfield, et al. [74]
				Souza-Melo et al. [75]
				Ogungbe et al. [58]
	Glyoxalase II	2P1E [76]	1.90	Ogungbe et al. [58]
		2P18 [76]	1.80	
	Nicotinamidase	3R2J [59]	2.68	
	Thiol-dependent reductase I	4AGS [77]	2.30	
	Trypanothione Reductase	4APN [78]	3.20	Shah, et al. [80] Ogungbe et al. [58]
		4APN [78]	3.20	
		2YAU [79]	2.50	
4ADW [78]		3.61		
<i>L. major</i>	Cathepsin B	Homologue model	-	Ogungbe et al. [58]

	Cysteine synthase	4AIR [55]	1.80	Bernal, Coy-Barrera.[57]
	Deoxyuridine triphosphate nucleotidohydrolase	2YAY [81]	1.86	Ogungbe et al. [58]
		2YB0 [81]	2.28	
	Dihydroorotate dehydrogenase	3GYE [82]	2.00	
		3MJY [83]	1.96	
		3MHU [62]	1.85	
	Leishmanolysin - GP63	1LML [84]	1.86	Shah, et al [80].
	Methionyl-tRNA synthetase	3KFL [61]	2.00	Ogungbe et al. [58]
	N-myristoyltransferase NMT	2WSA [85]	1.60	
		3H5Z [85]	1.49	
		4A30 [86]	1.50	
	Nucleoside diphosphate kinase b	3NGS [87]	1.80	
		3NGT [87]	2.57	
		3NGU [87]	2.29	
	Nucleoside hydrolase,	1EZR [88]	2.50	
	Oligopeptidase B,	2XE4 [89]	1.65	
	Phosphodiesterase 1	2R8Q [90]	1.50	
	Pteridine reductase 1	1E7W [91]	1.75	Ogungbe et al. [58]
		1W0C [92]	2.60	Ogungbe et al. [58]
		2BF7 [93]	2.40	Bernal, Coy-Barrera. [57]
		3H4V [54]	2.40	
		2QHX [54]	2.61	
	Trypanothione synthetase	2VOB [56]	2.30	Ogungbe et al. [58]
	Tyrosyl-tRNA synthetase,	3P0H [94]	3.00	
		3P0J [94]	2.89	Ogungbe et al. [58]
	Uridine diphosphate- glucose pyrophosphorylase	2OEF [60]	2.40	
		2OEG [60]	2.30	
<i>L. mexicana</i>	Glyceraldehyde-3- phosphate dehydrogenase	1A7K [95]	2.80	Ogungbe et al. [58]
		1GYP [96]	2.80	
	Glycerol-3-phosphate	1EVZ [97]	2.80	

	dehydrogenase	1M66 [98]	1.90	
		1N1E [63]	1.90	
		1N1G [98]	2.50	
	Phosphoglucose isomerase,	1Q50 [99]	2.60	
		1T10 [99]	2.35	
	Pyruvate kinase,	1PKL [100]	2.35	
		3HQP [101]	2.30	
		3PP7 [102]	2.35	
	Phosphomannomutase,	2I54 [103]	2.10	
		2I55 [103]	2.90	
	Triosephosphate isomerase	2VXN [104]	0.82	
		2Y61 [105]	0.99	
<i>L. panamensis</i>	Leishmanolysin - GP63	Homologue model	-	Mercado-Camargo, et al. [106]

Based on the results of the study, using the crystal structure of *L. infantum* CYP51 (PDB ID: 3L4D) [64], Warfield et al. used molecular docking calculations performed in Molegro Virtual docker (v.5.0. Molegro ApS, Aarhus, Denmark) to test a series of antiparasitic sterol-like compounds and their structural congeners, attempting to identify potential antiprotozoal drugs [74]. Some of these compounds can be found in Asteraceae species, especially in the *Taraxacum* genus [107-109]. Docking calculations were validated using ketoconazole and docking calculations were validated using ketoconazole and N-1-(2,4-dichlorophenyl)-2-(1H-imidazol-1-yl)ethyl)-4-(5-phenyl-1,3,4-oxadiazol-2-yl)benzamide, which showed redocking scores of −148.25 kcal/mol and −128.63 kcal/mol, respectively [74].

Interestingly, *epi*-oleanolic acid (compound **25**) and its derivatives presented high docking affinities for *L. infantum* CYP51 compared with the other sterols evaluated in this study, with a docking score of −105.7 kcal/mol. These docking results showed that *epi*-oleanolic acid derivatives with hydroxycinnamoyl groups interacted more at the active site of CYP51, establishing a hydrogen bond (H-bond) with A290 and steric interactions with the heme-cofactor. The best-ranked derivatives also interacted with L355, M459, F48, V356, and M357, a critical amino acid that interacts with lanosterol. Therefore, the inhibitory capacity of compound **25** and its derivatives was validated experimentally against CYP51. This study represents an interesting starting point for the identification of new therapies

against *Leishmania*, however only shows promising docking scores and further studies are necessary [74].

Oleanolic acid (compound **26**, 3 β -hydroxyolean-12-en-28-oic acid), a pentacyclic triterpenoid compound that has been isolated from several species of the Asteraceae family, including *Aspilia africana*, *Taraxacum officinale*, *Calendula officinalis*, and *Baccharis uncinella*, among others [110-112]. Souza Melo et al., using *L. infantum* CYP51 (PDB ID: 3L4D) [74], evaluated the antileishmanial activity of compound **26**, using promastigotes and amastigotes from three *Leishmania* species: *L. braziliensis*, *L. amazonensis*, and *L. infantum*. They reported IC₅₀ values ranging from 30.5 μ M to 68.8 μ M, with low cytotoxicity against mouse peritoneal macrophages. Molecular docking calculations were performed to establish the possible mechanism of action for compound **26** against CYP51 protein and to establish the molecular properties of this triterpenoid [75].

Molecular property calculations were studied in Spartan '14 software. Similar values were observed between compound **26** and lanosterol for van der Waals volume and the area; however, compound **26** showed a higher hydrophobic character (CLogP: 8.17) compared with lanosterol (CLogP: 7.10) and three times the polar surface area (PSA), with 49.241 $\text{\AA}^2$ compared with 17.950 $\text{\AA}^2$ for lanosterol. The energetic calculations suggested that compound **26** contains an additional region that is not present in lanosterol that is capable of interacting strongly with a hypothetical target [75].

Molecular docking was performed using GOLD software (v 4.1, CCDC Software Limited), and the crystal structure of *L. infantum* CYP51 (PDB ID: 3L4D) was obtained with fluconazole (PDB ID: TPF) as the ligand [64]. All of these *in silico* calculations were compared with those for lanosterol, a triterpenoid that is structurally like compound **26** and is a known, natural CYP51 substrate. For lanosterol, a strong interaction was observed with M357, in addition to the hydrophobic pocket, interacting with the residues Y102, M105, F109, V113, T115, and M283. Additionally, the methyl moiety of carbon-14 in lanosterol interacts with the heme-cofactor. In contrast, compound **26** did not establish any interactions with M357 or any residues of the hydrophobic pocket. The pentacyclic system of compound **26** establishes a similar spatial disposition in the pocket as the cyclopentanoperhydrophenanthrene in lanosterol. Finally, the authors highlighted the coordination of the carboxyl group in compound **26** (which has a similar orientation as the methyl moiety of carbon-14 in lanosterol) with the heme-cofactor. From these results, compound **26** shows leishmanicidal activity with low cytotoxicity against mouse peritoneal

macrophages might be a good candidate for the development of new leishmanicidal drugs, being established that acts against CYP51 through a different method than lanosterol [75]. β -sitosterol (compound **23**) is a phytosterol (a steroid subgroup) that has presented multiple pharmacological activities and is distributed in diverse genera of Asteraceae, such as *Achillea*, *Cichorium*, and *Aspilia*, among others [71,113,114].

Shah et al. isolated compound **23** from *Ifloga spicata* and evaluated its *in vitro* anti-leishmanial activity against *L. tropica* promastigotes. Compound **23** showed an IC₅₀ value of 22.2 μ M, whereas the control drug (glucantime) presented an IC₅₀ value of 14.6 μ M. Molecular docking calculations were performed in MOE using the crystal structures of *L. major* leishmanolysin (GP63, PDB: 1LML) [84] and *L. infantum* trypanothione reductase (TR, PDB ID: 4APN) [78] receptors [80].

The docking results showed that compound **23** interacted with TR and GP63, with binding energies of –61.54 kcal/mol and –33.24 kcal/mol, respectively. No redocking results were reported. The presence of the hydroxyl group in compound 23 may be responsible for the activity of the compound at the active site of TR, as the OH moiety of compound 23 establishes two H-bonds with G16 and A159, and this same OH group interacts at the active site of GP63 with D134 to establish a unique H-bond. Additionally, an H- π interaction with T226 was also observed [69]. Although these results are promising for compound **23**, proteins for *L. tropica* were not used in this study, and similarity results between the TR and GP63 structures for *L. tropica* were not reported.

Recently, Mercado-Camargo et al., using the same GP63 target (PDB: 1LML) as a template and built an *L. panamensis* GP63 homolog model using the SWISS-MODEL web server [115]. A series of biflavonoids were tested using AutoDock Vina in PyRx 0.8 [116]. *In vitro* tests were not performed. Among the 24 biflavonoids, quercitrin (compound **27**) was identified, which has been isolated from the Asteraceae genera *Solidago* and *Tagetes* [117]. Compound **27** shows moderate docking values of –8.2 kcal/mol for *L. major* and –7.5 kcal/mol for the homolog model of *L. panamensis* GP63. Amphotericin B was used as a control, which achieved docking values lower than –10.7 kcal/mol for GP63 in both species. Biflavonoids from the *Lanariaceae* and *Podocarpaceae* were the best-docked molecules in this study [106].

4. Conclusions

In recent years, cheminformatics tools have been widely used to identify safer and more effective treatments against *Leishmania*, which has included but was not limited to

the identification of new hits, the modification of existing molecules, and to better understand the parasitic life cycle. Vast numbers of secondary metabolites from Asteraceae have demonstrated antiparasitic activities, both *in vitro* and *in vivo*. Asteraceae, due to its broad distribution worldwide, represents an interesting source of potential new drug candidates against *Leishmania*.

Several studies that have used *in silico* methodologies to study compounds from Asteraceae have been published. Therefore, strengthening the specialized DBs for Asteraceae compounds represents an important step for better classifying and systematizing the large diversity of chemicals found in this family, to facilitate the performance of new computational structure-based and ligand-based virtual screening studies that allow a great number of molecules to be processed, with the aim of discovering new drug candidates against *Leishmania* at lower costs and with fewer side effects.

However, from this review, we identified the absence of *in vitro* results to verify the potential predicted activity in computational studies for secondary Asteraceae metabolites. The development of these assays is essential to advance in the search for new chemotherapies against the *Leishmania* parasites. This work presented the current panorama regarding this issue and seeks to be a starting point for future studies.

References

- [1].Gherbi, R.; Bounechada, M.; Latrofa, M.S.; Annoscia, G.; Tarallo, V.D.; Dantas-Torres, F.; Otranto, D. Phlebotomine sand flies and *Leishmania* species in a focus of cutaneous leishmaniasis in Algeria. *PLoS Negl Trop Dis* **2020**, *14*, e0008024, doi:10.1371/journal.pntd.0008024.
- [2].Akhoundi, M.; Kuhls, K.; Cannet, A.; Votýpka, J.; Marty, P.; Delaunay, P.; Sereno, D. A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Trop. Dis.* **2016**, *10*, e0004349.
- [3].Organization, W.H. Leishmaniasis. Available online: (accessed on 11 July 2020).
- [4].Herrera, G.; Barragan, N.; Luna, N.; Martinez, D.; De Martino, F.; Medina, J.; Nino, S.; Paez, L.; Ramirez, A.; Vega, L.; et al. An interactive database of *Leishmania* species distribution in the Americas. *Sci Data* **2020**, *7*, 110, doi:10.1038/s41597-020-0451-5.
- [5].Santos, S.S.; de Araujo, R.V.; Giarolla, J.; Seoud, O.E.; Ferreira, E.I. Searching for drugs for Chagas disease, leishmaniasis and schistosomiasis: a review. *Int J Antimicrob Agents* **2020**, *55*, 105906, doi:10.1016/j.ijantimicag.2020.105906.
- [6].DNDi. Towards a new generation of treatments for Leishmaniasis. **2018**.

- [7].Weng, H.B.; Chen, H.X.; Wang, M.W. Innovation in neglected tropical disease drug discovery and development. *Infect Dis Poverty* **2018**, *7*, 67, doi:10.1186/s40249-018-0444-1.
- [8].Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M.D.F.; Scotti, M.T. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. *Molecules* **2017**, *22*, 79.
- [9].Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770-803.
- [10].Rohloff, J.; Hymete, A.; Tariku, Y. Chapter 11 - Plant-Derived Natural Products for the Treatment of Leishmaniasis. In *Studies in Natural Products Chemistry*, Atta ur, R., Ed.; Elsevier: 2013; Volume 39, pp. 381-429.
- [11].Cheong, D.H.; Tan, W.D.; Wong, W.F.; Tran, T.J.P.R. Anti-malarial drug, artemisinin and its derivatives for the treatment of respiratory diseases. *Pharmacol. Res.* **2020**, 104901.
- [12].Tu, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med* **2011**, *17*, 1217-1220, doi:10.1038/nm.2471.
- [13].Ouellette, M.; Drummelsmith, J.; Papadopoulou, B. Leishmaniasis: drugs in the clinic, resistance and new developments. *Drug Resist Updat* **2004**, *7*, 257-266, doi:10.1016/j.drug.2004.07.002.
- [14].Aït-Oudhia, K.; Gazanion, E.; Vergnes, B.; Oury, B.; Sereno, D. Leishmania antimony resistance: what we know what we can learn from the field. *Mol. Microbiol.* **2011**, *109*, 1225.
- [15].Croft, S.; Oliaro, P. Leishmaniasis chemotherapy—challenges and opportunities. *Clin. Microbiol. Infect.* **2011**, *17*, 1478-1483.
- [16].Croft, S.L.; Yardley, V. Chemotherapy of leishmaniasis. *Curr. Pharm. Des.* **2002**, *8*, 319-342.
- [17].Yu, W.; MacKerell, A.D., Jr. Computer-Aided Drug Design Methods. *Methods Mol Biol* **2017**, *1520*, 85-106, doi:10.1007/978-1-4939-6634-9_5.
- [18].Patel, H.M.; Noolvi, M.N.; Sharma, P.; Jaiswal, V.; Bansal, S.; Lohan, S.; Kumar, S.S.; Abbot, V.; Dhiman, S.; Bhardwaj, V. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Medicinal Chemistry Research* **2014**, *23*, 4991-5007, doi:10.1007/s00044-014-1072-3.
- [19].Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys Rev* **2017**, *9*, 91-102, doi:10.1007/s12551-016-0247-1.

- [20].Zhang, D.-C.; Xing, Y.-P.; Xu, L.; Zhao, R.; Yang, Y.-Y.; Zhang, T.-T.; Li, S.-N.; Kang, T.-G. The complete mitochondrial genome of *Arctium lappa* (Campanulales, Asteraceae). *Mitochondrial DNA Part B* **2020**, *5*, 1722-1723.
- [21].Facility, G.G.B.I. Asteraceae in Species 2000 & ITIS Catalogue of Life: 2019, Catalogue of Life. Available online: <https://www.gbif.org/species/3065> (accessed on 15 july 2020).
- [22].Scotti, M.T.; Herrera-Acevedo, C.; Oliveira, T.B.; Costa, R.P.O.; Santos, S.; Rodrigues, R.P.; Scotti, L.; Da-Costa, F.B. Sistemax, an Online Web-Based Cheminformatics Tool for Data Management of Secondary Metabolites. *Molecules* **2018**, *23*, 103, doi:10.3390/molecules23010103.
- [23].Acevedo, C.H.; Scotti, L.; Scotti, M.T. In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018**, *13*, 634-645, doi:10.1002/cmdc.201700743.
- [24].Blunt, J.W.; Munro, M.H.G. 22 Is There an Ideal Database for Natural Products Research? In *Osborn, A.; Goss, R.J.; Carter, G.T. Natural Products: Discourse, Diversity, and Design. John Wiley & Sons, Inc; Wiley Online Library: 2014; pp. 413-431.*
- [25].Press, C. CRC Dictionary of Natural Products. Available online: http://dnp.chemnetbase.com/HelpFiles/DNP_Introduction.pdf (accessed on 13 july 2020).
- [26].Valli, M.; dos Santos, R.N.; Figueira, L.D.; Nakajima, C.H.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Development of a natural products database from the biodiversity of Brazil. *J Nat Prod* **2013**, *76*, 439-444, doi:10.1021/np3006875.
- [27].Banerjee, P.; Erehman, J.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II—a database of natural products. *Nucleic Acids Res.* **2015**, *43*, D935-D939.
- [28].Loub, W.D.; Farnsworth, N.R.; Soejarto, D.D.; Quinn, M.L. NAPRALERT: computer handling of natural product research data. *J Chem Inf Comput Sci* **1985**, *25*, 99-103, doi:10.1021/ci00046a009.
- [29].Afendi, F.M.; Okada, T.; Yamazaki, M.; Hirai-Morita, A.; Nakamura, Y.; Nakamura, K.; Ikeda, S.; Takahashi, H.; Altaf-Ul-Amin, M.; Darusman, L.K. KNApSACk family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant Cell Physiol* **2012**, *53*, e1-e1.
- [30].Sorokina, M.; Steinbeck, C. Review on natural products databases: where to find data in 2020. *J. Cheminf.* **2020**, *12*, 1-51.

- [31].Füllbeck, M.; Michalsky, E.; Dunkel, M.; Preissner, R. Natural products: sources and databases. *Nat. Prod. Rep.* **2006**, *23*, 347-356.
- [32].de Souza, M.C.; Rosa, A.L.; Poschenrieder, C.; Tolrà, R.; Da Costa, F.B. Fingerprinting metabolomics in tropical mistletoes: A case study with facultative aluminum-accumulating species. *Phytochem. Lett.* **2018**, *25*, 90-94.
- [33].AsterDB. AsterBioChem in House Database. Available online: <http://www.asterbiochem.org/asterdb> (accessed on 14 july 2020).
- [34].Raj, S.; Sasidharan, S.; Balaji, S.; Dubey, V.K.; Saudagar, P.J.J.o.P.; Proteomics. Review on natural products as an alternative to contemporary anti-leishmanial therapeutics. *J. Proteins Proteomics* **2020**, *11*, 135–158.
- [35].Rocha, L.G.; Almeida, J.R.; Macedo, R.O.; Barbosa-Filho, J.M. A review of natural products with antileishmanial activity. *Phytomedicine* **2005**, *12*, 514-535, doi:10.1016/j.phymed.2003.10.006.
- [36].Cavalcanti, É.B.; de Paulo Emerenciano, V.; Scotti, L.; Scotti, M. Flavonoids From Asteraceae as Multitarget Source of Compounds Against Protozoal Diseases. In *Multi-Scale Approaches in Drug Discovery*; Elsevier: 2017; pp. 149-190.
- [37].Emerenciano, V.; Militão, J.; Campos, C.; Romoff, P.; Kaplan, M.; Zambon, M.; Brant, A. Flavonoids as chemotaxonomic markers for Asteraceae. *Biochem. Syst. Ecol.* **2001**, *29*, 947-957.
- [38].Moraes Neto, R.N.; Setúbal, R.F.B.; Higino, T.M.M.; Brelaz-de-Castro, M.C.A.; da Silva, L.C.N.; Aliança, A.S.d.S. Asteraceae plants as sources of compounds against Leishmaniasis and Chagas disease. *Front. Pharmacol.* **2019**, *10*, 477.
- [39].Lowis, D.R. HQSAR: a new, highly predictive QSAR technique. *Tripes Technical Notes* **1997**, *1*, 17.
- [40].Trossini, G.H.; Maltarollo, V.G.; Schmidt, T.J. Hologram QSAR studies of antiprotozoal activities of sesquiterpene lactones. *Molecules* **2014**, *19*, 10546-10562, doi:10.3390/molecules190710546.
- [41].Schmidt, T.J.; Nour, A.M.; Khalid, S.A.; Kaiser, M.; Brun, R. Quantitative structure–antiprotozoal activity relationships of sesquiterpene lactones. *Molecules* **2009**, *14*, 2062-2076.
- [42].Zhang, D.; Haruna, M.; McPhail, A.T.; Lee, K.-H. Cytotoxic germacranolides of *Elephantopus carolinianus* and the structure and stereochemistry of isodeoxyelephantopin. *Phytochemistry* **1986**, *25*, 899-904.

- [43].Sosa, A.M.; Amaya, S.; Salamanca Capusiri, E.; Gilabert, M.; Bardon, A.; Gimenez, A.; Vera, N.R.; Borkosky, S.A. Active sesquiterpene lactones against *Leishmania amazonensis* and *Leishmania braziliensis*. *Nat Prod Res* **2016**, *30*, 2611-2615, doi:10.1080/14786419.2015.1126260.
- [44].Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* **2011**, *32*, 1466-1474, doi:10.1002/jcc.21707.
- [45].Tasdemir, D.; Kaiser, M.; Brun, R.; Yardley, V.; Schmidt, T.J.; Tosun, F.; Rüedi, P. Antitrypanosomal and antileishmanial activities of flavonoids and their analogues: in vitro, in vivo, structure-activity relationship, and quantitative structure-activity relationship studies. *Antimicrob. Agents Chemother.* **2006**, *50*, 1352-1364.
- [46].Panda, S.K.; Luyten, W. Antiparasitic activity in Asteraceae with special attention to ethnobotanical use by the tribes of Odisha, India. *Parasite* **2018**, *25*.
- [47].Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem* **2008**, *8*, 1555-1572, doi:10.2174/156802608786786624.
- [48].Chibli, L.A.; Rosa, A.L.; Nonato, M.C.; Da Costa, F.B. Untargeted LC–MS metabolomic studies of Asteraceae species to discover inhibitors of *Leishmania major* dihydroorotate dehydrogenase. *Metabolomics* **2019**, *15*, 59.
- [49].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.
- [50].Reis, R.A.G.; Calil, F.A.; Feliciano, P.R.; Pinheiro, M.P.; Nonato, M.C. The dihydroorotate dehydrogenases: Past and present. *Arch Biochem Biophys* **2017**, *632*, 175-191, doi:10.1016/j.abb.2017.06.019.
- [51].Herrera-Acevedo, C.; Maia, M.D.S.; Cavalcanti, É.B.V.S.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Selection of antileishmanial sesquiterpene lactones from Sistemax database using a combined ligand-/structure-based virtual screening approach. *Molecular Diversity* **2020**, 1-17.
- [52].Brannigan, J.A.; Smith, B.A.; Yu, Z.; Brzozowski, A.M.; Hodgkinson, M.R.; Maroof, A.; Price, H.P.; Meier, F.; Leatherbarrow, R.J.; Tate, E.W.; et al. N-myristoyltransferase from *Leishmania donovani*: structural and functional characterisation of a potential drug target for visceral leishmaniasis. *J Mol Biol* **2010**, *396*, 985-999, doi:10.1016/j.jmb.2009.12.032.

- [53].Dufe, V.T.; Ingner, D.; Heby, O.; Khomutov, A.R.; Persson, L.; Al-Karadaghi, S. A structural insight into the inhibition of human and *Leishmania donovani* ornithine decarboxylases by 1-amino-oxy-3-aminopropane. *Biochemical Journal* **2007**, *405*, 261-268.
- [54].Cavazzuti, A.; Paglietti, G.; Hunter, W.N.; Gamarro, F.; Piras, S.; Loriga, M.; Allecca, S.; Corona, P.; McLuskey, K.; Tulloch, L. Discovery of potent pteridine reductase inhibitors to guide antiparasite drug development. *Proc Natl Acad Sci USA* **2008**, *105*, 1448-1453.
- [55].Fyfe, P.K.; Westrop, G.D.; Ramos, T.; Muller, S.; Coombs, G.H.; Hunter, W.N. Structure of *Leishmania* major cysteine synthase. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2012**, *68*, 738-743, doi:10.1107/S1744309112019124.
- [56].Fyfe, P.K.; Oza, S.L.; Fairlamb, A.H.; Hunter, W.N. *Leishmania* trypanothione synthetase-amidase structure reveals a basis for regulation of conflicting synthetic and hydrolytic activities. *J Biol Chem* **2008**, *283*, 17672-17680, doi:10.1074/jbc.M801850200.
- [57].Bernal, F.A.; Coy-Barrera, E. In-silico analyses of sesquiterpene-related compounds on selected *Leishmania* enzyme-based targets. *Molecules* **2014**, *19*, 5550-5569, doi:10.3390/molecules19055550.
- [58].Ogungbe, I.V.; Setzer, W.N. In-silico *Leishmania* target selectivity of antiparasitic terpenoids. *Molecules* **2013**, *18*, 7761-7847, doi:10.3390/molecules18077761.
- [59].Gazanion, E.; Garcia, D.; Silvestre, R.; Gerard, C.; Guichou, J.; Labesse, G.; Seveno, M.; Cordeiro - Da - Silva, A.; Ouaisi, A.; Sereno, D. The *Leishmania* nicotinamidase is essential for NAD⁺ production and parasite proliferation. *Mol. Microbiol.* **2011**, *82*, 21-38.
- [60].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.
- [61].Larson, E.T.; Kim, J.E.; Zucker, F.H.; Kelley, A.; Mueller, N.; Napuli, A.J.; Verlinde, C.L.; Fan, E.; Buckner, F.S.; Van Voorhis, W.C.; et al. Structure of *Leishmania* major methionyl-tRNA synthetase in complex with intermediate products methionyladenylate and pyrophosphate. *Biochimie* **2011**, *93*, 570-582, doi:10.1016/j.biochi.2010.11.015.
- [62].Cordeiro, A.T.; Feliciano, P.R.; Pinheiro, M.P.; Nonato, M.C. Crystal structure of dihydroorotate dehydrogenase from *Leishmania* major. *Biochimie* **2012**, *94*, 1739-1748.

- [63].Choe, J.; Guerra, D.; Michels, P.A.; Hol, W.G. Leishmania mexicana glycerol-3-phosphate dehydrogenase showed conformational changes upon binding a bi-substrate adduct. *J. Mol. Biol.* **2003**, *329*, 335-349.
- [64].Hargrove, T.Y.; Wawrzak, Z.; Liu, J.; Nes, W.D.; Waterman, M.R.; Lepesheva, G.I. Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14 α -demethylase (CYP51) from Leishmania infantum. *J. Biol. Chem.* **2011**, *286*, 26838-26848.
- [65].Carrillo-Hormaza, L.; Mora, C.; Alvarez, R.; Alzate, F.; Osorio, E. Chemical composition and antibacterial activity against Enterobacter cloacae of essential oils from Asteraceae species growing in the Páramos of Colombia. *Ind. Crops Prod.* **2015**, *77*, 108-115.
- [66].Maggio, A.; Rosselli, S.; Bruno, M.; Spadaro, V.; Raimondo, F.M.; Senatore, F. Chemical composition of essential oil from Italian populations of Artemisia alba Turra (Asteraceae). *Molecules* **2012**, *17*, 10232-10241, doi:10.3390/molecules170910232.
- [67].Saleh, M.A.; Belal, M.H.; el-Baroty, G. Fungicidal activity of Artemisia herba alba Asso (Asteraceae). *J Environ Sci Health B* **2006**, *41*, 237-244, doi:10.1080/03601230500354774.
- [68].Singh, O.; Khanam, Z.; Misra, N.; Srivastava, M.K. Chamomile (Matricaria chamomilla L.): An overview. *Pharmacogn Rev* **2011**, *5*, 82-95, doi:10.4103/0973-7847.79103.
- [69].Calderon, J.S.; Quijano, L.; Gómez-Garibay, F.; Sanchez, D.M.; Rios, T.; Fronczek, F.R. Sesquiterpene lactones from Stevia ovata and crystal structure of 11, 13-dehydroeriolin. *Phytochemistry* **1987**, *26*, 1747-1750.
- [70].Sebisubi, F.M.; Odyek, O.; Anokbonggo, W.W.; Ogwal-Okeng, J.; Carcache-Blanco, E.J.; Ma, C.; Orjala, J.; Tan, G.T. Antimalarial activity of Aspilia pruliseta, a medicinal plant from Uganda. *Planta Med* **2010**, *76*, 1870-1873, doi:10.1055/s-0030-1250028.
- [71].Gomes, A.; Saha, A.; Chatterjee, I.; Chakravarty, A.K. Viper and cobra venom neutralization by beta-sitosterol and stigmasterol isolated from the root extract of Pluchea indica Less. (Asteraceae). *Phytomedicine* **2007**, *14*, 637-643, doi:10.1016/j.phymed.2006.12.020.
- [72].Venugopal, V.; Sen, B.; Datta, A.K.; Banerjee, R. Structure of cyclophilin from Leishmania donovani at 1.97 Å resolution. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2007**, *63*, 60-64.

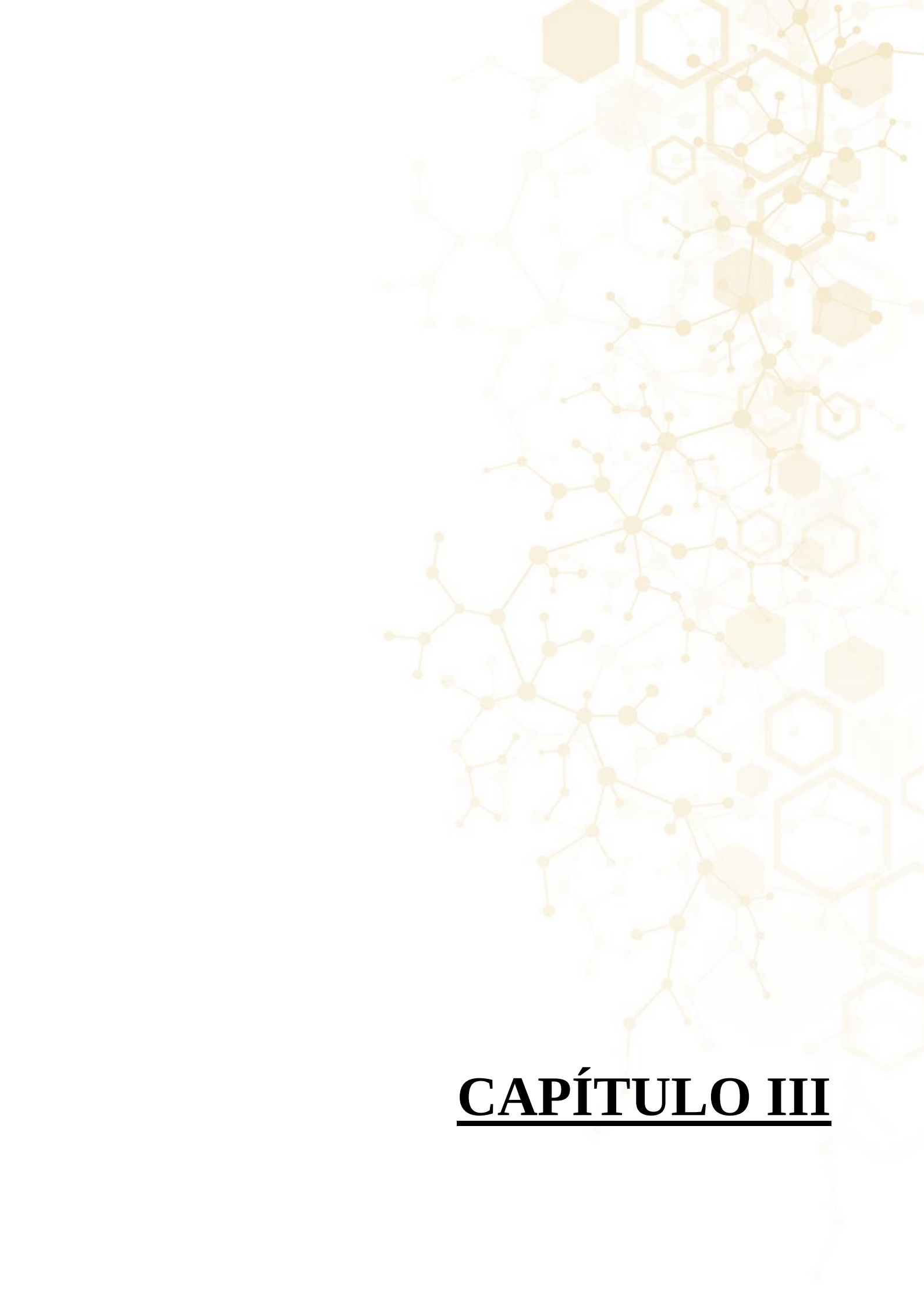
- [73].Venugopal, V.; Datta, A.K.; Bhattacharyya, D.; Dasgupta, D.; Banerjee, R. Structure of cyclophilin from *Leishmania donovani* bound to cyclosporin at 2.6 Å resolution: correlation between structure and thermodynamic data. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 1187-1195.
- [74].Warfield, J.; Setzer, W.N.; Ogungbe, I.V. Interactions of antiparasitic sterols with sterol 14 α -demethylase (CYP51) of human pathogens. *SpringerPlus* **2014**, *3*, 679.
- [75].Melo, T.S.; Gattass, C.R.; Soares, D.C.; Cunha, M.R.; Ferreira, C.; Tavares, M.T.; Saraiva, E.; Parise-Filho, R.; Braden, H.; Delorenzi, J.C. Oleanolic acid (OA) as an antileishmanial agent: Biological evaluation and in silico mechanistic insights. *Parasitol. Int.* **2016**, *65*, 227-237.
- [76].Sousa Silva, M.; Barata, L.; Ferreira, A.E.; Romão, S.; Tomás, A.M.; Ponces Freire, A.; Cordeiro, C. Catalysis and structural properties of *Leishmania infantum* glyoxalase II: trypanothione specificity and phylogeny. *Biochemistry* **2008**, *47*, 195-204.
- [77].Fyfe, P.K.; Westrop, G.D.; Silva, A.M.; Coombs, G.H.; Hunter, W.N. *Leishmania* TDR1 structure, a unique trimeric glutathione transferase capable of deglutathionylation and antimonial prodrug activation. *Proc Natl Acad Sci U S A* **2012**, *109*, 11693-11698, doi:10.1073/pnas.1202593109.
- [78].Baiocco, P.; Poce, G.; Alfonso, S.; Cocozza, M.; Porretta, G.C.; Colotti, G.; Biava, M.; Moraca, F.; Botta, M.; Yardley, V.; et al. Inhibition of *Leishmania infantum* trypanothione reductase by azole-based compounds: a comparative analysis with its physiological substrate by X-ray crystallography. *ChemMedChem* **2013**, *8*, 1175-1183, doi:10.1002/cmdc.201300176.
- [79].Ilari, A.; Baiocco, P.; Messori, L.; Fiorillo, A.; Boffi, A.; Gramiccia, M.; Di Muccio, T.; Colotti, G. A gold-containing drug against parasitic polyamine metabolism: the X-ray structure of trypanothione reductase from *Leishmania infantum* in complex with auranofin reveals a dual mechanism of enzyme inhibition. *Amino Acids* **2012**, *42*, 803-811, doi:10.1007/s00726-011-0997-9.
- [80].Shah, S.M.; Ullah, F.; Ayaz, M.; Sadiq, A.; Hussain, S.; Shah, S.A.A.; Nadhman, A. β -Sitosterol from *Ifloga spicata* (Forssk.) Sch. Bip. as potential anti-leishmanial agent against *leishmania tropica*: docking and molecular insights. *J Steroids* **2019**, *148*, 56-62.
- [81].Hemsworth, G.R.; Moroz, O.V.; Fogg, M.J.; Scott, B.; Bosch-Navarrete, C.; Gonzalez-Pacanowska, D.; Wilson, K.S. The crystal structure of the *Leishmania* major deoxyuridine triphosphate nucleotidohydrolase in complex with nucleotide analogues,

- dUMP, and deoxyuridine. *J Biol Chem* **2011**, 286, 16470-16481, doi:10.1074/jbc.M111.224873.
- [82].Cordeiro, A.; Feliciano, P.; Nonato, M. Crystal structure of Leishmania major dihydroorotate dehydrogenase. *Biochimie* **2012**.
- [83].Cheleski, J.; Rocha, J.R.; Pinheiro, M.P.; Wiggers, H.J.; da Silva, A.B.; Nonato, M.C.; Montanari, C.A. Novel insights for dihydroorotate dehydrogenase class 1A inhibitors discovery. *Eur. J. Med. Chem.* **2010**, 45, 5899-5909.
- [84].Schlagenhauf, E.; Etges, R.; Metcalf, P. The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63). *Structure* **1998**, 6, 1035-1046, doi:10.1016/s0969-2126(98)00104-x.
- [85].Frearson, J.A.; Brand, S.; McElroy, S.P.; Cleghorn, L.A.T.; Smid, O.; Stojanovski, L.; Price, H.P.; Guthrie, M.L.S.; Torrie, L.S.; Robinson, D.A. N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature* **2010**, 464, 728-732.
- [86].Brand, S.; Cleghorn, L.A.T.; McElroy, S.P.; Robinson, D.A.; Smith, V.C.; Hallyburton, I.; Harrison, J.R.; Norcross, N.R.; Spinks, D.; Bayliss, T. Discovery of a novel class of orally active trypanocidal N-myristoyltransferase inhibitors. *Journal of medicinal chemistry* **2012**, 55, 140-152.
- [87].Souza, T.A.C.B.; Trindade, D.M.; Tonoli, C.C.C.; Santos, C.R.; Ward, R.J.; Arni, R.K.; Oliveira, A.H.C.; Murakami, M.T. Molecular adaptability of nucleoside diphosphate kinase b from trypanosomatid parasites: stability, oligomerization and structural determinants of nucleotide binding. *Molecular BioSystems* **2011**, 7, 2189-2195.
- [88].Shi, W.; Schramm, V.L.; Almo, S.C. Nucleoside hydrolase from Leishmania major Cloning, expression, catalytic properties, transition state inhibitors, and the 2.5-Å crystal structure. *Journal of Biological Chemistry* **1999**, 274, 21114-21120.
- [89].McLuskey, K.; Paterson, N.G.; Bland, N.D.; Isaacs, N.W.; Mottram, J.C. Crystal structure of Leishmania major oligopeptidase B gives insight into the enzymatic properties of a trypanosomatid virulence factor. *Journal of biological chemistry* **2010**, 285, 39249-39259.
- [90].Wang, H.; Yan, Z.; Geng, J.; Kunz, S.; Seebeck, T.; Ke, H. Crystal structure of the Leishmania major phosphodiesterase LmjPDEB1 and insight into the design of the parasite - selective inhibitors. *Molecular microbiology* **2007**, 66, 1029-1038.
- [91].Gourley, D.G.; Schüttelkopf, A.W.; Leonard, G.A.; Luba, J.; Hardy, L.W.; Beverley, S.M.; Hunter, W.N. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. *Nature structural biology* **2001**, 8, 521-525.

- [92].McLuskey, K.; Gibellini, F.; Carvalho, P.; Avery, M.A.; Hunter, W.N. Inhibition of *Leishmania major* pteridine reductase by 2, 4, 6-triaminoquinazoline: structure of the NADPH ternary complex. *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60*, 1780-1785.
- [93].Schüttelkopf, A.W.; Hardy, L.W.; Beverley, S.M.; Hunter, W.N. Structures of *Leishmania major* pteridine reductase complexes reveal the active site features important for ligand binding and to guide inhibitor design. *Journal of molecular biology* **2005**, *352*, 105-116.
- [94].Larson, E.T.; Kim, J.E.; Castaneda, L.J.; Napuli, A.J.; Zhang, Z.; Fan, E.; Zucker, F.H.; Verlinde, C.L.M.J.; Buckner, F.S.; Van Voorhis, W.C. The double-length tyrosyl-tRNA synthetase from the eukaryote *Leishmania major* forms an intrinsically asymmetric pseudo-dimer. *Journal of molecular biology* **2011**, *409*, 159-176.
- [95].Kim, H.; Hol, W.G.J. Crystal structure of *Leishmania mexicana* glycosomal glyceraldehyde-3-phosphate dehydrogenase in a new crystal form confirms the putative physiological active site structure. *Journal of molecular biology* **1998**, *278*, 5-11.
- [96].Kim, H.; Feil, I.K.; Verlinde, C.L.M.J.; Petra, P.H.; Hol, W.G.J. Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry* **1995**, *34*, 14975-14986.
- [97].Suresh, S.; Turley, S.; Oppendoes, F.R.; Michels, P.A.M.; Hol, W.G.J. A potential target enzyme for trypanocidal drugs revealed by the crystal structure of NAD-dependent glycerol-3-phosphate dehydrogenase from *Leishmania mexicana*. *Structure* **2000**, *8*, 541-552.
- [98].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. *Chemistry & biology* **2002**, *9*, 1189-1197.
- [99].Cordeiro, A.T.; Michels, P.A.M.; Delboni, L.F.; Thiemann, O.H. The crystal structure of glucose - 6 - phosphate isomerase from *Leishmania mexicana* reveals novel active site features. *European journal of biochemistry* **2004**, *271*, 2765-2772.
- [100].Rigden, D.J.; Phillips, S.E.V.; Michels, P.A.M.; Fothergill-Gilmore, L.A. The structure of pyruvate kinase from *Leishmania mexicana* reveals details of the allosteric transition and unusual effector specificity. *Journal of molecular biology* **1999**, *291*, 615-635.

- [101].Morgan, H.P.; McNae, I.W.; Nowicki, M.W.; Hannaert, V.; Michels, P.A.M.; Fothergill-Gilmore, L.A.; Walkinshaw, M.D. Allosteric mechanism of pyruvate kinase from *Leishmania mexicana* uses a rock and lock model. *Journal of Biological Chemistry* **2010**, *285*, 12892-12898.
- [102].Morgan, H.P.; McNae, I.W.; Nowicki, M.W.; Zhong, W.; Michels, P.A.M.; Auld, D.S.; Fothergill-Gilmore, L.A.; Walkinshaw, M.D. The trypanocidal drug suramin and other trypan blue mimetics are inhibitors of pyruvate kinases and bind to the adenosine site. *Journal of Biological Chemistry* **2011**, *286*, 31232-31240.
- [103].Kedzierski, L.; Malby, R.L.; Smith, B.J.; Perugini, M.A.; Hodder, A.N.; Ilg, T.; Colman, P.M.; Handman, E. Structure of *Leishmania mexicana* phosphomannomutase highlights similarities with human isoforms. *Journal of molecular biology* **2006**, *363*, 215-227.
- [104].Alahuhta, M.; Wierenga, R.K. Atomic resolution crystallography of a complex of triosephosphate isomerase with a reaction - intermediate analog: New insight in the proton transfer reaction mechanism. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1878-1888.
- [105].Venkatesan, R.; Alahuhta, M.; Pihko, P.M.; Wierenga, R.K. High resolution crystal structures of triosephosphate isomerase complexed with its suicide inhibitors: The conformational flexibility of the catalytic glutamate in its closed, liganded active site. *Protein Science* **2011**, *20*, 1387-1397.
- [106].Mercado-Camargo, J.; Cervantes-Ceballos, L.; Vivas-Reyes, R.; Pedretti, A.; Serrano-Garcia, M.L.; Gomez-Estrada, H. Homology Modeling of Leishmanolysin (gp63) from *Leishmania panamensis* and Molecular Docking of Flavonoids. *ACS Omega* **2020**, *5*, 14741-14749, doi:10.1021/acsomega.0c01584.
- [107].Pütter, K.M.; van Deenen, N.; Müller, B.; Fuchs, L.; Vorwerk, K.; Unland, K.; Bröker, J.N.; Scherer, E.; Huber, C.; Eisenreich, W. The enzymes OSC1 and CYP716A263 produce a high variety of triterpenoids in the latex of *Taraxacum koksaghyz*. *Sci. Rep.* **2019**, *9*, 1-13.
- [108].Lim, T. *Taraxacum officinale*. In *Edible Medicinal And Non-Medicinal Plants*; Springer: 2014; pp. 516-536.
- [109].Furuno, T.; Kamiyama, A.; AKASHI, T.; Usui, M.; TAKAHASHI, T.; AYABE, S.-i. Triterpenoid constituents of tissue cultures and regenerated organs of *Taraxacum officinale*. *Plant Tissue Cult. Lett.* **1993**, *10*, 275-280.

- [110].Pollier, J.; Goossens, A. Oleanolic acid. *Phytochemistry* **2012**, *77*, 10-15, doi:10.1016/j.phytochem.2011.12.022.
- [111].Ayeleso, T.B.; Matumba, M.G.; Mukwevho, E. Oleanolic acid and its derivatives: biological activities and therapeutic potential in chronic diseases. *Molecules* **2017**, *22*, 1915.
- [112].Passero, L.F.; Bonfim-Melo, A.; Corbett, C.E.; Laurenti, M.D.; Toyama, M.H.; de Toyama, D.O.; Romoff, P.; Favero, O.A.; dos Grecco, S.S.; Zalewsky, C.A.; et al. Anti-leishmanial effects of purified compounds from aerial parts of *Baccharis uncinella* C. DC. (Asteraceae). *Parasitol Res* **2011**, *108*, 529-536, doi:10.1007/s00436-010-2091-8.
- [113].Saeidnia, S.; Manayi, A.; Gohari, A.R.; Abdollahi, M. The story of beta-sitosterol-a review. *Eur. J. Med. Plants* **2014**, 590-609.
- [114].El-Lakany, A.M.; Aboul-Ela, M.A.; Abdul-Ghani, M.M.; Mekky, H. Chemical constituents and biological activities of *Cichorium intybus* L. *Nat. Prod. Sci.* **2004**, *10*, 69-73.
- [115].Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **2018**, *46*, W296-W303, doi:10.1093/nar/gky427.
- [116].Dallakyan, S.; Olson, A.J. Small-molecule library screening by docking with PyRx. In *Chemical biology*; Springer: 2015; pp. 243-250.
- [117].Bohm, B.A.; Stuessy, T.F. *Flavonoids of the sunflower family (Asteraceae)*; Springer Science & Business Media: 2001.

The background of the page is a light cream color, overlaid with a complex, abstract pattern of thin, golden-brown lines. These lines form a network of interconnected nodes and clusters, resembling a molecular structure or a complex graph. Interspersed within this network are numerous hexagonal shapes, some of which are solid golden-brown and others that are hollow outlines. The overall effect is a sophisticated, scientific, or technological aesthetic.

CAPÍTULO III

A leishmaniose visceral (LV) também conhecida como Calazar, é uma doença causada por *L. infantum* e *L. donovani*, sendo a segunda doença tropical e subtropical mais letal e a sétima em perda de anos de vida ajustados por incapacidade [1]. É transmitido pela picada de um flebotomíneo infectado e pode afetar pessoas de todas as idades, embora em áreas endêmicas sua incidência seja maior em crianças devido à imunidade adquirida por adultos [2].

Os parasitas *Leishmania* são organismos dimórficos que vivem e se replicam no intestino dos flebótomos na forma flagelada (promastigota) ou como formas aflageladas (amastigotas) nas células de mamíferos. No hospedeiro mamífero, esses parasitas preferencialmente infectam células fagocíticas, principalmente macrófagos e células dendríticas [3]. Cada uma dessas formas parasitárias apresenta características únicas que desempenham um papel crucial no ciclo de vida e na patogenicidade da *Leishmania donovani*.

Na busca por novas quimioterapias contra o parasita *Leishmania*, as sesquiterpenlactonas, metabólitos secundários vegetais com ocorrência especialmente generalizada na família de plantas Compositae/Asteraceae, emergiram como compostos promissores [4]. A Asteraceae é uma das famílias que se destaca na busca por novas moléculas bioativas, especialmente aquelas com atividade antiparasitária [5].

Especificamente, respeito a *Leishmania donovani*, Schmidt et al. relataram as bioatividades *in vitro* de 40 sesquiterpenlactonas (SLs). Duas SLs de *Xanthium brasiliicum*, 4,15-dinor-1,11(13)-xanthadiene-3,5 β :12,8 β -diolide e 8-epixanthatin 1 β ,5 β -epoxide, foram as mais seletivas [6]. Em outro estudo, três SLs de *Anthemis auriculata* apresentaram atividade contra amastigotas de *L. donovani*, com valores de IC₅₀ de 3,27, 8,18 e 12,5 mg/mL, respectivamente [7].

No entanto, tenham sido identificados várias SLs promissórias, a maior parte destes estudos relatados estão focados em um único alvo. Recentemente, o conceito de *hits* multialvo ganhou relevância na pesquisa de doenças parasitárias como a leishmaniose. O desenvolvimento de medicamentos que podem interagir simultaneamente com vários alvos é uma abordagem promissora para o tratamento de doenças complexas. Em comparação com o uso de combinações de medicamentos de alvo único, os medicamentos multialvo têm vantagens em termos de maior eficácia, perfil de segurança aprimorado e administração mais simples [8].

Assim, no presente capítulo, utilizando métodos de triagem virtual em lactonas sesquiterpênicas (SLs), o estudo empregou abordagens baseadas em ligantes e em

estrutura, incluindo a análise de proteínas-alvo, para buscar compostos multitarget contra quatro proteínas de *Leishmania donovani*. Uma análise de consenso identificou 13 SLs potenciais contra a leishmaniose a partir do banco de dados Sistemax, mostrando agentes promissórios direcionados a múltiplas enzimas de *L. donovani*. O estudo demonstra uma abordagem estratégica para a descoberta de medicamentos no combate à leishmaniose.

. REFERÊNCIAS

- [1]. COSTA, Carlos HN et al. From Infection to Death: An Overview of the Pathogenesis of Visceral Leishmaniasis. **Pathogens**, v. 12, n. 7, p. 969, 2023.
- [2]. SCARPINI, Sara et al. Visceral leishmaniasis: epidemiology, diagnosis, and treatment regimens in different geographical areas with a focus on pediatrics. **Microorganisms**, v. 10, n. 10, p. 1887, 2022.
- [3]. KIMA, Peter E. The amastigote forms of Leishmania are experts at exploiting host cell processes to establish infection and persist. **International journal for parasitology**, v. 37, n. 10, p. 1087-1096, 2007.
- [4]. PAULSEN, Evy. The sesquiterpene lactone mix: A review of past, present and future aspects. **Contact Dermatitis**, 2023.
- [5]. LAURELLA, Laura C. et al. Antiprotozoal compounds from Mikania periplocifolia (Asteraceae). **Fitoterapia**, v. 167, p. 105499, 2023.
- [6]. SCHMIDT, Thomas J. et al. Quantitative Structure– Antiprotozoal Activity Relationships of Sesquiterpene Lactones. **Molecules**, v. 14, n. 6, p. 2062-2076, 2009.
- [7]. KARIOTI, Anastasia et al. Trypanocidal, leishmanicidal and cytotoxic effects of anthecotulide-type linear sesquiterpene lactones from Anthemis auriculata. **Phytomedicine**, v. 16, n. 8, p. 783-787, 2009.
- [8]. ZHANG, Weilin; PEI, Jianfeng; LAI, Luhua. Computational multitarget drug design. **Journal of chemical information and modeling**, v. 57, n. 3, p. 403-412, 2017.

Selection of Antileishmanial Sesquiterpene Lactones from SistematX Database using a combined ligand/structure-based virtual screening approach.

Chonny Herrera-Acevedo^{1,2}, Mayara Dos Santos Maia¹, Élide Batista Vieira Sousa Cavalcanti¹, Ericsson Coy-Barrera², Luciana Scotti¹, Marcus Tullius Scotti^{1,*}

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil.

² Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, 250247, Cajicá, Colombia

*Contact author. Email: mtscotti@gmail.com

Artigo publicado em Molecular Diversity (IF: 2.943)

DOI Number: <https://doi.org/10.1007/s11030-020-10139-6>

Ano de publicação: 2020

Abstract: Leishmaniasis refers to a complex of diseases, caused by the intracellular parasitic protozoans belonging to the genus *Leishmania*. Among the three types of disease manifestations, the most severe type is visceral leishmaniasis, which is caused by *Leishmania donovani*, and is diagnosed in more than 20,000 cases annually, worldwide. Because the current therapeutic options for disease treatment are associated with several limitations, the identification of new potential leads/drugs remains necessary. In this study, a combined approach was used, based on two different virtual screening (VS) methods, which were designed to select promising antileishmanial agents from among the entire sesquiterpene lactone (SL) dataset registered in SistematX, a web interface for managing a secondary metabolite database that is accessible by multiple platforms on the Internet. Thus, a ChEMBL dataset, including 3,159 and 1,569 structures that were previously tested against *Leishmania donovani* amastigotes and promastigotes *in vitro*, respectively, was used to develop two random forest models, which performed with greater than 74% accuracy in both the cross-validation and test sets. Subsequently, a ligand-based virtual screening assay was performed against the 1,306 SistematX-registered SLs. In parallel, the crystal structures of three *Leishmania donovani* target proteins, N-myristoyltransferase, ornithine decarboxylase, and mitogen-activated protein kinase 3, and a homology model of

pteridine reductase 1 were used to perform a structure-based virtual screening, using molecular docking, of the entire Sistemax SL dataset. The consensus analysis of these two virtual screening approaches resulted in the normalization of probability scores and identified 13 promising, enzyme-targeting, antileishmanial SLs from Sistemax that may act against *Leishmania donovani*.

Keywords: *Leishmania donovani*; Sesquiterpene lactones; Ligand-based virtual screening; Structure-based virtual screening; Machine learning; Sistemax database.

1. Introduction

Leishmaniasis refers to a complex of diseases, caused by the intracellular parasitic protozoans of the genus *Leishmania*, a representative of the order Kinetoplastida and the family Trypanosomatidae [1]. These organisms are heteroxenous parasites that require two hosts to complete their life cycle, a vertebrate and an invertebrate, and commonly infect the hematophagous dipterans, more commonly known as sandflies, belonging to the genus *Phlebotomus*, of the order Diptera in the subfamily Phlebotominae (family Psychodidae) [2,3]. Sandflies become infected when they bite an infected individual, ingesting host-infected macrophages or free amastigotes from the blood or tissues. Upon reaching the insect's midgut, amastigotes develop into promastigotes. These flagellar forms, after rapid multiplication, become infective, and migratory promastigotes are regurgitated and introduced into the skin of the next host when the insect takes a new blood meal [4].

The clinical manifestation of leishmaniasis depends on the complexity of the interaction between the host's immune system and the protozoan type, with four recognized disease presentations: cutaneous, cutaneous mucosal, diffuse cutaneous, and, visceral leishmaniasis (VL), which is the most severe [5,6]. The number of VL cases, annually, is estimated at 20,000, worldwide, caused by *Leishmania donovani*, and VL can be fatal without treatment [7,8]. VL is a chronic, systematic disease, with marked clinical manifestations, including fever, hepatomegaly, splenomegaly, cutaneous/mucosal pallor, diarrhea, and weight loss [9]. In addition, a canine form of visceral-cutaneous leishmaniasis develops a clinical manifestation that resembles the presentation in humans [10].

Currently available therapeutic drugs are associated with prolonged treatment times and intense side-effects, which often result in patients abandoning treatment; therefore, the identification of new drugs or lead structures is urgently necessary; however, leishmaniasis

is considered to be a *Neglected Tropical Disease*, due to a lack of research and the poor development of new drugs over many decades [11-14]. Many classes of chemicals, including natural products, have provided interesting leads for the treatment of parasites, particularly sesquiterpene lactones (SLs) [11,15]. Several studies have evaluated the leishmanicidal activity of SLs as potential anti-*Leishmania* drugs [16]. Schmidt et al. reported the *in vitro* bioactivities of 40 SLs against *L. donovani* and other parasites and their cytotoxicities against L6 rat skeletal myoblasts. Two SLs, 4,15-dinor-1,11(13)-xanthadiene-3,5 β :12,8 β -diolide and 8-epixanthatin 1 β ,5 β -epoxide, which were isolated from *Xanthium brasiliicum*, were identified as the most selective SLs against *L. donovani* [17]. In another study, three irregular, linear SLs derived from *Anthemis auriculata*, 4-hydroxyanthecotulide, anthecotulide, and 4-acetoxyanthecotulide, showed activity against *L. donovani* amastigotes, with half-maximal inhibitory concentration (IC₅₀) values of 3.27, 8.18, and 12.5 mg/mL, respectively [18].

Enzymes and metabolites that are present in the parasite but absent from their mammalian hosts are considered ideal targets for rational drug design [19]. Several potential enzymatic targets have been explored during the development of new leishmaniasis treatment drugs, which aim to utilize natural products as enzymatic inhibitors. Ornithine decarboxylase (ODC, E.C. 4.1.1.17) is a key enzyme involved in polyamine biosynthesis in *L. donovani*, catalyzing the conversion of ornithine to putrescine, which is a precursor of other polyamines. The inhibition of ODC depletes the parasite of trypanothione, causing an imbalance in redox metabolism and increasing reactive oxygen species [20-23].

N- myristoyltransferase (NMT, E.C. 2.3.1.97) is a ubiquitous enzyme that catalyzes the attachment of myristic acid (a 14-carbon saturated fatty acid) to the amino-terminal glycine residue of a subset of eukaryotic proteins [24,25]. Mitogen-activated protein (MAP) kinase 3 (MPK3, E.C. 2.7.11.25) is a component of signal transduction pathways and an important regulator of cell differentiation and cell proliferation in eukaryotic cells. Thus far, ten MAP kinases have been identified in *Leishmania mexicana* [26]. Pteridine reductase 1 (PTR1, E.C. 1.5.1.33), an NADPH-dependent short-chain reductase, is responsible for the unusual salvage of pterin in *Leishmania* and acts as a metabolic bypass for drugs that target dihydrofolate reductase [27].

In this study, two virtual screening (VS) approaches were utilized to examine the potential activity of 1,306 SLs against *L. donovani*, from the compounds registered in the SistematX database. Initially, a ligand-based VS approach was developed, using random

forest (RF) models constructed against the two parasitic forms, amastigotes and trypomastigotes. Subsequently, using a homolog model of *L. donovani* PTR1 and the crystal structures of three enzymes expressed by this *Leishmania* species, a structure-based VS approach was also applied to the Sistemax SL dataset. Finally, a combined approach, using both VS approaches, was performed to identify SLs with potential activity against *L. donovani*, and to establish potential mechanisms of action for the identified compounds.

2. Material and methods

2.1. Database

From the ChEMBL database (<https://www.ebi.ac.uk/chembl/>), we selected a diverse set of structures that were initially classified according to their predicted activity against the two *L. donovani* parasitic forms: amastigote (5,500 structures) and promastigote (2,045 structures). These compounds were classified according to pIC₅₀ values [$-\log\text{IC}_{50}$ (mol/L)]; therefore, we stratified them into active (pIC₅₀ $\geq$ 4.7) and inactive (pIC₅₀ < 4.7) structures. Due to the variability of experimental protocols used by the ChEMBL database, a qualitative pattern was used, to partially minimize the differences in activity values associated with different experimental protocols and strains. IC₅₀ values represent the concentration required to inhibit 50% of parasite growth. For each ChEMBL dataset, 35 amastigote-associated and six promastigote-associated SL structures, which were not included in the dataset used for the ligand-based VS, and their respective pIC₅₀ values were added, to increase the representativeness of the generated models with regard to the chemical space of this class of secondary metabolites.

The applicability domain (APD), based on Euclidean distances, was used to identify those compounds in the test set for which predictions may be unreliable; compounds were considered unreliable if they had APD values higher than $d + Z\sigma$, where d was the average Euclidean distance and σ was the standard deviation of the set of samples in the training set with lower than average Euclidean distance values relative to all samples in the training set. The parameter Z is an empirical cut off value, and 0.5 was used as the default value [28]. Structures with pIC₅₀ values ranging from 4.6 to 4.7 (range of 0.1 units) were excluded, to avoid edge effects and improve the predictive capacity of the models. Excluding these structures minimized the differences in activity values resulting from errors and differences in experimental protocols [29].

Data curation of the datasets was performed, according to the suggested procedures in the literature [30-32]. Standardizer software [Jchem, version 16.11.28 (2016),

calculation module developed by ChemAxon, <http://www.chemaxon.com/>] was used to canonize all simplified molecular-input line-entry system (SMILES) codes. After duplicate structures were removed, those with higher pIC₅₀ values were eliminated. The use of only those compounds with lower activity values facilitated the generation of more restrictive models.

After dataset curation, 3,159 structures for amastigotes (1,564 active and 1,595 inactive) and 1,569 structures for promastigotes (756 active and 813 inactive) were included in the analysis. An SL dataset was obtained from the Sistemax database (<http://sistemax.ufpb.br>), and a total of 1,306 molecules from this dataset were used in this study.

For predictions, those SL structures that were also included in the ChEMBL dataset were excluded for each parasitic form. For all structures, SMILES codes were used as the input data in Marvin [ChemAxon, version 16.11.28 (2016), calculation module developed by ChemAxon, <http://www.chemaxon.com/>]. We used standardizer software [Jchem, version 16.11.28 (2016), calculation module developed by ChemAxon, <http://www.chemaxon.com/>]. ChemAxon was used to canonize the structures, add hydrogens, perform aromatic form conversions, and clean molecular graphs in three dimensions. This software was used to generate and optimize conformers for the initial structure (represented by the root node in the tree). Those molecules that presented structural problems during the three-dimensional (3D) generation were manually corrected using Marvin.

2.2. Volsurf+ descriptors

The 3D structures of the identified molecules, in special data file (SDF) format, were used as input data in Volsurf+, v. 1.0.7 [33] and were subjected to molecular interaction fields (MIFs), to generate descriptors, using the following probes: N₁ (amide nitrogen–hydrogen-bond donor probe), O (carbonyl oxygen–hydrogen-bond acceptor probe), OH₂ (water probe) and DRY (hydrophobic probe). Additional non-MIF-derived descriptors were generated, resulting in a total of 128 descriptors [33]. One of the main advantages of using VolSurf descriptors is the relatively low influence of conformational sampling and averaging on these descriptors [34].

2.3 RF models

Knime 3.1.0 software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org) [35] was used to perform all of the following analyses. Initially, the descriptors calculated in the Volsurf+ program were imported, in comma-separated value (CSV) format, and the “Partitioning” node in the stratified sampling option was used to classify 80% of the initial dataset as the training set and the remaining 20% as the test set. The model was generated by employing the modeling set and the RF algorithm, with a “5-fold external validation” procedure, using WEKA nodes. In the 5-fold cross-validation procedure, the dataset is divided five times into a modeling set (80%–20%).

After this modeling set (which was used to build and validate models) is divided additionally into multiple training (80%) and test sets (20%) [32,36]. The parameters selected for the RF models included the following: number of trees to build = 200; seed for random number generator = 1; and Gini Index, as a split criterion, for both the training and internal cross-validation sets.

From the confusion matrix, the internal and external performances of the selected models were analyzed, using the following parameters: sensitivity (true-positive rate), specificity (true-negative rate), and accuracy (overall predictability). In addition, to describe the true performance of the model with more clarity than can be obtained from accuracy alone, the receiver operating characteristic (ROC) curve was employed, using a “ROC curve” node, which uses the sensitivity and specificity parameters. The plotted ROC curve shows the true-positive (active) rate versus the false-positive rate (1 - specificity) [37].

In this representation, when a variable of interest cannot be distinguished between the two groups, the ROC area under the curve (AUC) value is 0.5, whereas the perfect separation between the values of the two groups, with no distribution overlap, results in a ROC AUC value of 1. Matthew’s correlation coefficient (MCC) was also calculated, wherein a value of 1 represents a perfect prediction, a value of 0 represents a random prediction, and a value of –1 represents total disagreement between the prediction and the observation [38].

2.4 False positive remover

To detect false-positive structures among the SLs that were classified as active in the two RF models, the substructure filter for the removal of pan assay interference compounds (PAINS) was used [39]. All SMILES codes for SLs classified as active were

submitted to PAINS removal (<http://www.cbligand.org/PAINS/>), and those structures that were classified as false-positives were excluded from the final analysis.

2.5 Homology model of *LdPTR1*

2.5.1. Homolog identification of target sequences, the selection of a protein template, and the alignment of the template and target sequences.

The target protein sequence was obtained from the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/pubmed>) [40]. The selection of a template protein was performed using the Basic Local Alignment Search Tool (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [28]. The RCSB Protein Data Bank (PDB, <https://www.rcsb.org/pdb/home/home.do>) [41] was used to obtain the protein structure. The template protein that was selected was *LmPTR1* (PDB ID: 1E7W) [42]. The alignment of multiple sequences was performed using FASTA (<http://www.ebi.ac.uk/Tools/sss/fasta/>), and the following values were obtained for the comparisons between *L. donovani* PTR1 (*LdPTR1*) and *Leishmania major* PTR1 (*LmPTR1*): 91.0% identity and 97.2% similarity

2.5.2. Construction and validation of the model

The *LdPTR1* model was constructed using the homology molecular modeling method, using MODELLER 9.18 software [43], which is based on spatial-constraint satisfaction modeling. Five models were generated, and the lowest energy model was chosen.

The stereochemical qualities of the model were evaluated with PROCHECK [44], which evaluated several stereochemical parameters, such as the torsional angles of the main chain, the torsional angles of the side chain, bad contacts or steric impediments, and planarity. PROCHECK generated a Ramachandran graph [45], which verified the allowed and unallowed regions of the main amino acid chain.

The structural quality was evaluated in VERIFY 3D software (<http://services.mbi.ucla.edu/SAVES/>), which analyzes the compatibility of the protein sequence with its 3D structure, according to the chemical environment, and WHAT IF (<http://swift.cmbi.ru.nl/servers/html/index.html>), which analyzes various structural parameters, such as the atomic contacts between residues. The software Discovery Studio Visualizer [46] was used to visualize the modeled protein.

2.6 Molecular docking

In addition to a homology model of PTR1, the structures of three *L. donovani* proteins, ODC (PDB ID: 2O00) [47], NMT (PDB ID: 2WUU) [48], and MPK3 (PDB ID: 402Z), in complex with their respective inhibitors: pyridoxal-5'-phosphate (PDB ID: PLP), 2-oxopentadecyl-CoA (PDB ID: NHW), and 5-[[6-(acetylamino)pyrimidin-4-yl]oxy}-N-{4-[(4-methylpiperazin-1-yl)methyl]-3(trifluoromethyl)phenyl}-2,3-dihydro-1H-indole-1-carboxamide (PDB ID: 046), were downloaded from PDB. Using Molegro 6.0.1 software, all water compounds were deleted from the enzyme structures, and the enzyme/compound structures were prepared using the same default parameter settings, in the same software package (Score function: MolDock Score; Ligand evaluation: Internal ES, Internal HBond, Sp2–Sp2 Torsions, all checked; Number of runs: 10 runs; Algorithm: MolDock SE; Maximum Interactions: 1500; Max. population size: 50; Max. steps: 300; Neighbor distance factor: 1.00; Max. number of poses returned: 5).

The docking procedure was performed using a grid, with a 15-Å radius and a 0.30-Å resolution, to cover the ligand-binding site for the four enzyme structures. All SMILES codes of the SLs classified as active in molecular docking calculations were submitted to PAINS removal (<http://www.cbligand.org/PAINS/>), and those structures that the filter classified as false-positive were excluded from the final analysis [36,39].

2.7 Molecular dynamics simulations

Protein-ligand complexes for the best-ranked molecules identified by the three VS approaches were each subjected to a 20-ns molecular dynamics (MD) simulation, using GROMACS 5.0, with the GROMOS 43a1 force field (FF) [49,50]. Using the webserver PROGRG26 topology generator (<http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg/submit.html>) [51], the topologies for the six analyzed ligands were generated. The same FF was used to prepare the topology of the homology model, *LdPTR1*. A 100-ps, restrained equilibration and a 1-ns MD simulation were performed, to refine the model prior to the evaluation of protein-ligand complexes.

The MD simulations were performed using an SPC water model of point load, extended in a dodecahedral box. Na⁺ and Cl⁻ ions were added to neutralize the overall charge on the system. The system was also balanced in 300 K, using the V-rescale algorithm, at 100 ps, represented by NVT (constant number of particles, volume, and

temperature), followed by equilibrium at 1 atm of pressure, using the Parrinello-Rahman algorithm of NPT (constant pressure, number of particles, and temperature), at 100 ps.

The root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration (RoG) plots were generated in Grace software (<http://plasma-gate.weizmann.ac.il/Grace/>), and the proteins and ligands were visualized using UCSF Chimera [52]. The molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method was used to calculate binding free energies, using the trajectories calculated by the MD simulations [53].

3. Results and discussion

Following the good practices established for quantitative structure-activity relationship (QSAR) and molecular docking studies [54,31,55], a combined ligand-based and structure-based VS study was performed, using a databank containing 1,306 SLs that was stored in Sistemax, a database developed at the Laboratory of Cheminformatics of the Federal University of Paraíba, which contains a wealth of useful information regarding natural products, including the locations of the species from which the indexed compounds were isolated [56]. Ligand-based VS was performed starting with a ChEMBL database (<https://www.ebi.ac.uk/chembl/>), which contained structures with previously demonstrated *in vitro* activity against *L. donovani*.

Data were curated from these datasets according to the procedures suggested in the literature [30,32]. Afterward, Volsurf+ descriptors (Volsurf+ program v. 1.0.7) [33,34] were calculated, using the 3D structures generated using ChemAxon JChem, v. 16.11.28 (2016), a calculation module developed by ChemAxon (<http://www.chemaxon.com/>). Using Knime 3.1.0 software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, <http://www.knime.org>) [35], two RF models were generated and validated for their abilities to determine the activity probabilities for the entire SL dataset (Figure 1).

In parallel, using the homologous protein *LdPTR1*, which based on the crystal structure of *L. major* PTR1 (*LmPTR1*) [42] and three crystal structures for potential target enzymes expressed by *L. donovani*, ODC [47], NMT [48] and MPK3, molecular docking was performed in Molegro virtual docker, 6.0.1 using the SL dataset. Finally, using the active probability values that were obtained from these two methodologies, a consensus analysis was performed to select those molecules with the best-combined values (Figure 1). MD simulations were also performed to refine and validate the *LdPTR1* homology model

and to evaluate the formation of protein-ligand complexes between *LdPTR1* and the best-ranked SLs, over time.

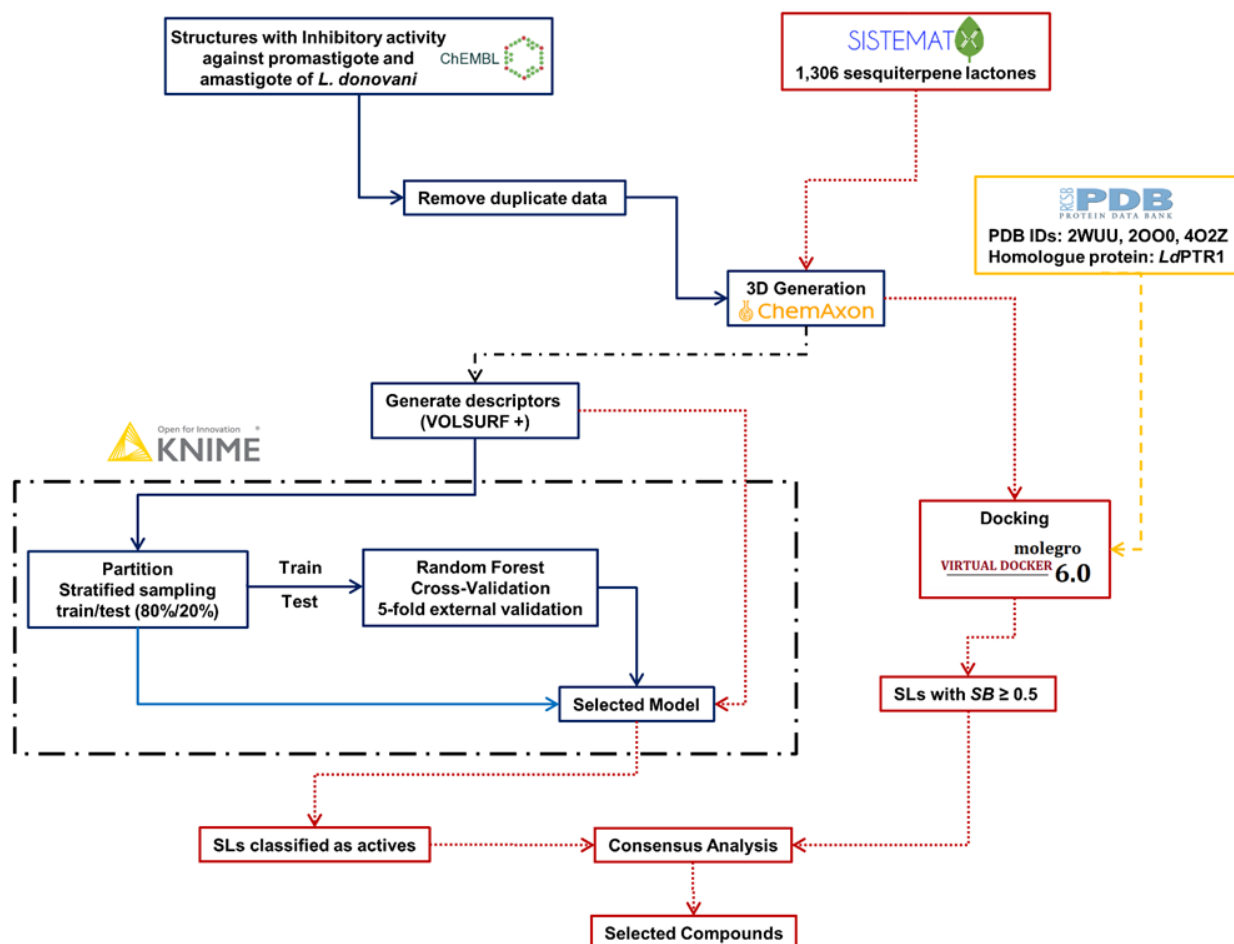


Figure 1. The VS methodology used in this study. Solid blue lines represent the two sets of compounds used to generate and validate the RF model for *L. donovani* amastigotes and promastigotes (clear blue line represents the external test set). The red dotted lines represent the SLs from Asteraceae, obtained from Sistemax (in-house database). The black dash-dot line represents both datasets (ChEMBL and Sistemax). The yellow dashed line represents the three *L. donovani* protein structures extracted from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB), (PDB ID: 2WUU, 2O00 and 4O2Z) and the homologous protein, *LdPTR1*. The dash-dot border delimits the process performed in KNIME software.

3.1 Ligand-based virtual screening

Volsurf+ descriptors (128) were calculated for two ChEMBL datasets: promastigote (1,628 structures) and amastigote (3,259 structures), which were classified as either active or inactive (binary classification), using a cut-off value of $\text{pIC}_{50} \geq 4.7$, which allowed the maximum representation of the chemical space for each class of structure (active and inactive). Subsequently, VolSurf+ descriptor values, together with their respective binary classifications, were used as input data, to generate two RF models in the Knime program. All 128 Volsurf+ descriptors were used to build the RF models. Several models were evaluated to minimize the false-positive rate of the models. Finally, for both models, 200 trees and the Gini Index, as a split criterion, were selected. Structures with pIC_{50} values between 4.6 and 4.7 (range of 0.1 units) were excluded, to avoid edge effects and improve the predictive capacity of the models, by minimizing potential activity differences due to errors and different experimental protocols. A 5-fold cross-validation procedure was performed, splitting the dataset five times into a modeling set (80%–20%). Only the modeling set, which was additionally divided into multiple training and test sets (80%–20%) was used to build and validate the models. [32,36].

Table 1. Summary of cross-validation and 5-fold external validation results, which were obtained using the RF algorithm on the total set of 3,159 compounds for *L. donovani* amastigotes, and 1,569 compounds for *L. donovani* promastigotes

Model	Outcome	Cross-validation	5-fold external validation
		(%)	(%)
Amastigote	Active	73.7	72.8
	Inactive	78.0	76.8
	Total	76.8	74.8
Promastigote	Active	72.1	74.8
	Inactive	78.9	77.3
	Total	75.6	76.1

For the training sets used in both RF models, the match percentage values approached 100%. For the cross-validation and test sets, values above 72.8% were obtained. In all cases, the inactive values (from 77.3% to 78.9%) were greater than the active values (from 72.1% to 74.8%), demonstrating that the obtained models were highly restrictive, which is

a fundamental requirement for this type of study, as restrictive models minimize the probability of obtaining false-positive structure selection and prevents inactive molecules from being predicted as active.

Sensitivity and specificity are two key parameters for the selection of RF models. Specificity, which is defined as the true-negative rate, was higher for promastigote models (78.9% and 77.3%) than for amastigote models (78.0% and 76.8%), for both the internal validation and test sets, respectively. Similarly, sensitivity, which is defined as the true-positive rate, showed that the percentage of predicted true-positive compounds was higher for the amastigote model (73.7%) than for the promastigote model, during cross-validation. For the test set, the promastigote model achieved a slightly higher specificity (74.8%) than the amastigote model.

The ROC curve is a quality parameter that plots the true-positive rate (sensitivity) against the false-positive rate (1 - specificity), and the values for the AUC of a ROC plot can range between 0 and 1 [37]. The maximum value is achieved when the perfect separation occurs between the values of two groups, whereas an AUC equal to 0.5 indicates that the variable of interest cannot be distinguished between the two groups (Figure 2, gray line).

The *L. donovani* RF models achieved AUC values greater than 0.82, demonstrating a high degree of differentiation between the active and inactive compounds in the ChEMBL dataset. AUC values for the amastigote model (Figure 2a), of 0.84 and 0.82 for the internal cross-validation and 5-fold external validation, respectively, were minimally lower than those for the promastigote model (Figure 2b), of 0.84 and 0.83. Additionally, MCC, which is a quality parameter, was determined for the two models, using all values in the confusion matrix, based on Equation 1.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (\text{Equation 1})$$

where TP is the true-positive rate, TN is the true-negative rate, FP is the false-positive rate, and FN is the false-negative rate [38]. An MCC value equal to 1 indicates a perfect correlation, a value of 0 indicates a random prediction, and a value of -1 indicates total disagreement between the prediction and the observation [38]. Here, similar MCC values were observed for both models, with a slightly higher MCC value for the amastigote model (0.52) in the 5-fold external validation. In contrast, the promastigote model achieved a major MCC value during cross-validation.

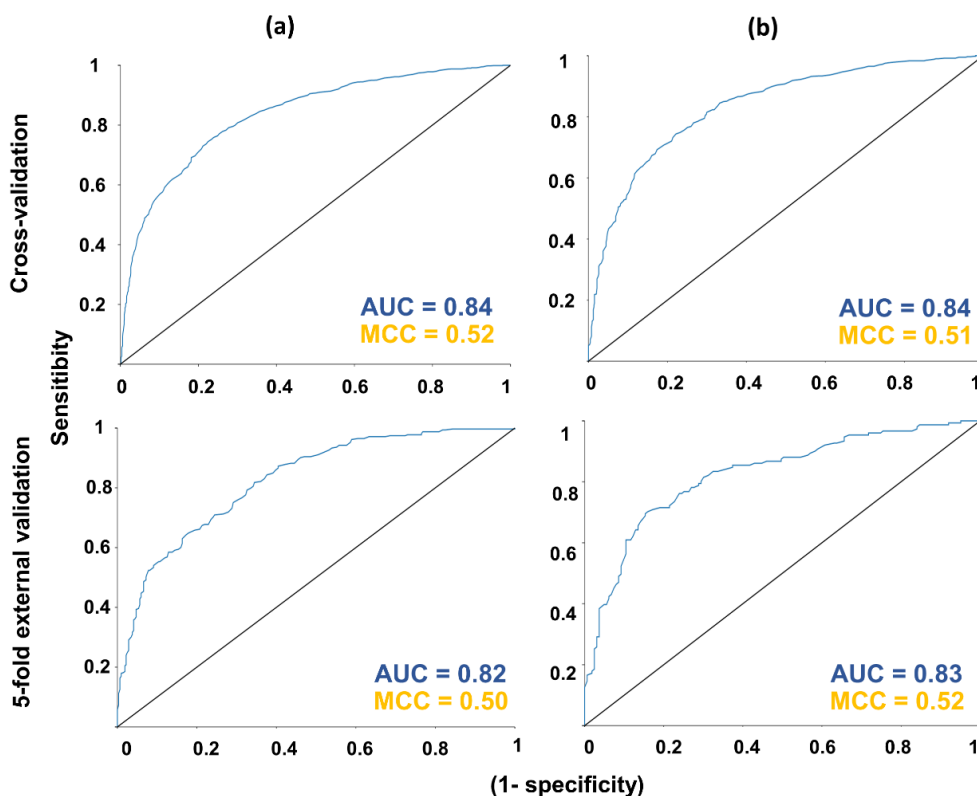


Figure 2. ROC plot, sensitivity versus (1 - specificity), generated by the selected RF model for cross-validation and 5-fold external validation of the **(a)** promastigote and **(b)** amastigote models. AUC = value of the Area Under the Curve; MCC = Matthews's Correlation Coefficient.

To identify any compounds in the test sets for both models and the SL dataset for which the predictions may be unreliable, the domain of applicability was calculated. Only one molecule, out of 630 structures in the test set, was classified as unreliable for the amastigote model. For the promastigote model, and for the SL dataset used for VS, all structures were classified as reliable.

Ligand-based VS was performed to predict which structures may potentially be active against the two parasitic forms of *Leishmania*, using the entire set of 1,306 SLs registered in SistematX. For the amastigote model, 712 molecules were classified as active, with active probability (**LB**) values ranging between 0.50 and 0.87. Two types of skeletons were observed among the five best-ranked structures: four pseudoguaianolides (Figure 3, structures **1** and **3-5**) and a germacranolide (structure **2**). Some specific structural features were observed among this group, specifically the presence of alkyl esters moieties in all hits, such as the two *cis*-related acetyl groups in the most active germacranolide and the most active SL, cumanin-diacetate (structure **1**), which is a secondary metabolite

commonly identified in plants from the *Ambrosia* genus (*A. psilostachya* and *A. acanthicarpa*) [57].

Similarly, an α -methylene- γ -lactone was observed in three of the five best-ranked structures (structures **1**, **3**, and **4**), the presence of this moiety has been associated with interactions between this type of metabolite and the sulfhydryl group of cysteine, through a Michael addition [58].

Interestingly, arnicolide C (structure **3**, Figure 3) and arnifolin (structure **4**), two SLs that are typically found in *Arnica montana* L., also known as mountain arnica [59], presented high **LB**-values during the ligand-based VS as two of the best-ranked molecules for the amastigote parasitic form. Structurally, these two molecules are pseudoguaianolides, with similar molecular weights, that only differ by the substitution of an isobutyrate group at C-6 in arnicolide C (structure **3**) and an angelate group (in addition to the presence of the α -methylene- γ -lactone system in C-11) in arnifolin (structure **4**).

For the promastigote model, only 14 of 1,306 screened SLs were classified as active. A different structural pattern was observed among the five best-ranked SLs for promastigotes compared with those identified for amastigotes. In all five cases, SLs with high molecular weights presented the highest **LB** values, including SL dimers of structures **1-4**. *Bedfordia* symmetric dimeric lactone (structure **6**, Figure 3), a secondary metabolite isolated from *Bedfordia salicina*, which is a species commonly found in Tasmania, reached an **LB** value of 0.66. This SL is composed of two eremophilanolide units, which are linked through a covalent bond between the carbon 8 of both structures (structure **7**, Figure 3).

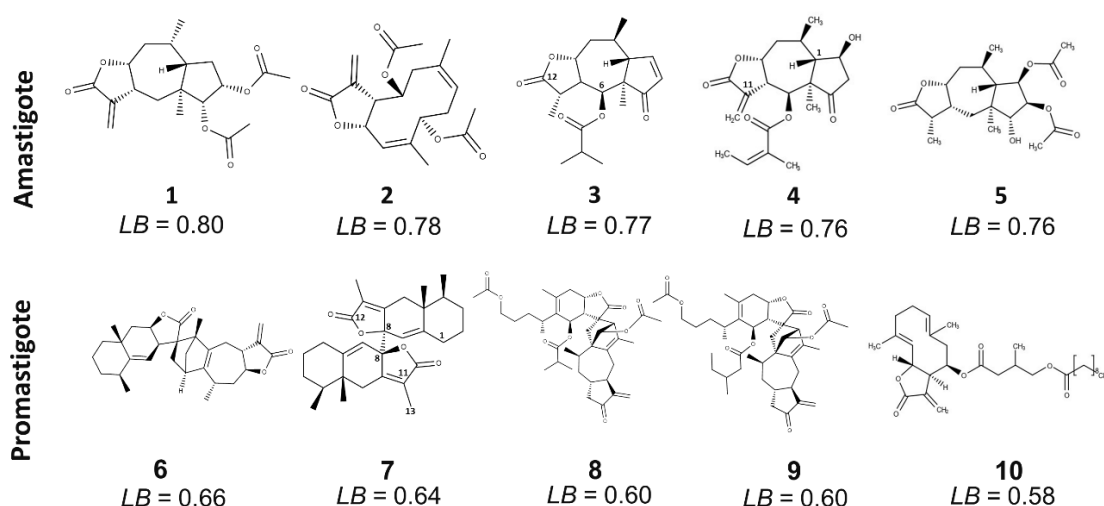


Figure 3. Potentially active sesquiterpene lactones (five best-ranked), identified using ligand-based VS, for the two-parasitic forms of *L. donovani*. **LB** = active probability value.

Similarly, carpedilactone G (structure **6**) also presented a high **LB** value against promastigotes and, similar to structure **7**, features an eremophilanolide skeleton. Additionally, two dimers isolated from *Inula japonica*, japonicone V (structure **8**) and japonicone U (structure **9**), which achieved **LB** values of 0.60, were classified by the ligand-based VS as potential antileishmanial structures against promastigotes. Previously, dimeric SLs from this plant have been reported to have anti-inflammatory and cytotoxic activities [60].

To examine the results for false-positive molecules, the entire set of 712 (amastigote) and 14 (promastigote) SLs that were classified as potentially active by the ligand-based VS were tested using the PAINS remover web tool [38]. All of the molecules that were classified as active against the promastigote parasitic form passed through the filter, whereas among the 712 molecules with **LB** values above 0.5 from the amastigote RF model, two structures, Neurolobatin B (**LB** = 0.508) and 3-chlorodehydroleucodin (**LB** = 0.544) did not pass the filter and were classified as false-positives.

3.2. Structure-based virtual screening

3.2.1. Homology model of *LdPTR1*

Molecular modeling was performed for the *L. donovani* PTR1 protein, based on homology with *L. major* PTR1 [42]. A good template for *LdPTR1* was obtained due to the high level of similarity between the target sequence (*LdPTR1*) and the template sequence (*LmPTR1*). To verify and validate the reliability and stereochemical qualities of the modeled protein, data from Ramachandran, VERIFY 3D, and WHAT IF graphs were considered. The Ramachandran plot showed that the main possible chain conformations included 88.3% with residues in the most favored regions, 11.7% with residues in allowed regions, and 0% outliers (Supplementary Figure S1). Because no residues were found in the outlier region, the generated model was considered satisfactory, and all residues in the active site were analyzed against the template sequence and found to be conserved [42]. The G factors, which indicate the quality of the covalent distance and the bond angle, were 0.15 for dihedrons and 0.09 for phi/psi. Positive, or non-negative, values indicate a model with good stereochemistry. According to the VERIFY 3D results, 85.07% of residues had mean 3D/1D scores ≥ 0.2 , which indicated a reliable model because more than 80% of amino acids had values of 0.2 in the 3D/1D profile. The quality of atomic contacts between the atoms of each residue was analyzed, using the Fine Packing Quality Control module of

WHAT IF, which compares the distribution of atom positions around each residue. The mean score of all wastes was -0.687. A score of less than -5.0 for a residue indicates poor or unusual atomic contacts. Finally, MD simulations (1 ns) were performed to refine and validate the homology model.

3.2.2. Molecular docking

In addition to *LdPTR1*, three *L. donovani* target proteins, whose crystal structures were obtained from the PDB databank, were analyzed by molecular docking: ODC (E.C. 4.1.1.17), NMT (E.C. 2.3.1.97), and MPK3 (E.C. 2.7.11.25). Initially, the methodology was validated by performing redocking with the ligand reported in the PDB crystal structure for each of the three *L. donovani* proteins used in this study. These values were not obtained for *LdPTR1*. The docking scores and their respective RMSD values are listed in Table 2.

Using the same parameters, a binding VS for the entire set of 1,306 SLs was performed. Based on the binding energy values, all tested molecules were ranked, using the following probability calculation.

$$SB = \frac{E_i}{E_{min}} \text{ if } E_i < E_{ligand} \text{ (Equation 2)}$$

where **SB** is the structure-based probability; E_i is the docking energy of compound i, where i ranges from 1 to 1,306 (SL dataset); E_{min} is the lowest energy value of the dataset; and E_{ligand} is the ligand energy from protein crystallography

This equation was proposed used to normalize the scores obtained from the molecular docking analyses during the structure-based VS and to generate values that were comparable to the active probability values obtained during the ligand-based VS. Structures that presented structure-based probability values (**SB**) above 0.5 were classified as active. An additional applied criterion was defined to select those structures with energies below that obtained for the ligand used in the crystallography study. This criterion was not applied to *LdPTR1* and, instead, SLs were only ranked according to the minimum docking energy value [61].

The analysis of the NMT protein identified 528 compounds with **SB** values above 0.5; however, only 490 of these had binding energy values lower than that for the PDB ligand, 2-oxopentadecyl-CoA (PDB ID: NHW), which was -50.4 KJ/mol. For MPK3 and ODC, the numbers of SLs with **SB** values > 0.5 were 243 and 8, respectively. For both

enzymes, only two structures (structures **13-14** for MPK3 and structures **14-15** for ODC) had lower energies than the respective inhibitors reported in the crystal structures. Finally, for *LdPTR1*, 346 molecules showed **SB** values greater than the cut-off value; however, no reference ligand was used for this homologous protein. To produce the probability of selecting molecules with high **SB** but without real activity against *Leishmania* (false-positives), all SLs that were classified as being active by the structure-based VS were filtered using the PAINS. For NMT only 1 of the 490 structures classified as active was identified as being a false-positive (Neurolobatin A); for PTR1 two structures with **SB** values higher than 0.5 were filtered out by PAINS (Neurolobatin A and B); for ODC and MPK3, no molecules were identified as being false-positive.

The best-ranked molecules for each enzyme are listed in Table 2 and Figure 4. All binding energy values can be found in the supporting information

Table 2. The docking energy (kJ/mol) for the best-ranked SLs, according to the structure-based approach, for each of the four *L. donovani* proteins studied. Ligand = docking energy (kJ/mol) for the PDB ligand.

<i>L. donovani</i> protein	SL-1	SL-2	Ligand	Redocking RMSD
NMT	11 (-98.8)	12 (-98.1)	(-50.4)	1.77
ODC	13 (-92.1)	14 (-83.7)	(-72.0)	0.14
MPK3	14 (-128.5)	15 (-114.4)	(-112.6)	0.29
<i>LdPTR1</i>	16 (-76.7)	17 (-76.3)	-	-

Guaianolide skeletons were predominant among the seven SLs that presented the highest **SB** values for the four enzymes studied, and structures **12**, **13**, **15**, and **17** (Figure 4) all belonged to this group of substructures. Similarly, two germacranolides, structures **11** and **16**, displayed very low docking energy values for NMT and *LdPTR1*, respectively. Structure **14**, an SL extracted from *Picradeniopsis woodhousei*, was classified as a disecoeudesmanolide and was the only compound identified by the structure-based VS associated with high scores for two *L. donovani* proteins (Table 2), MPK3 (−128.5 kJ/mol) and ODC (−83.7 kJ/mol).

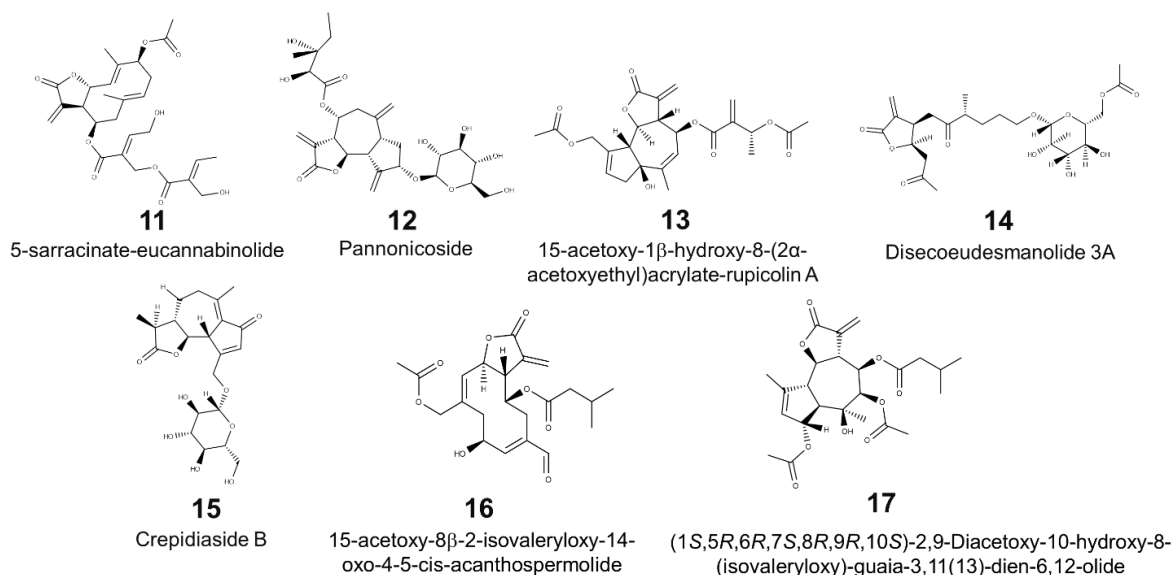


Figure 4. Best-ranked antileishmanial SLs from the structure-based virtual screening.

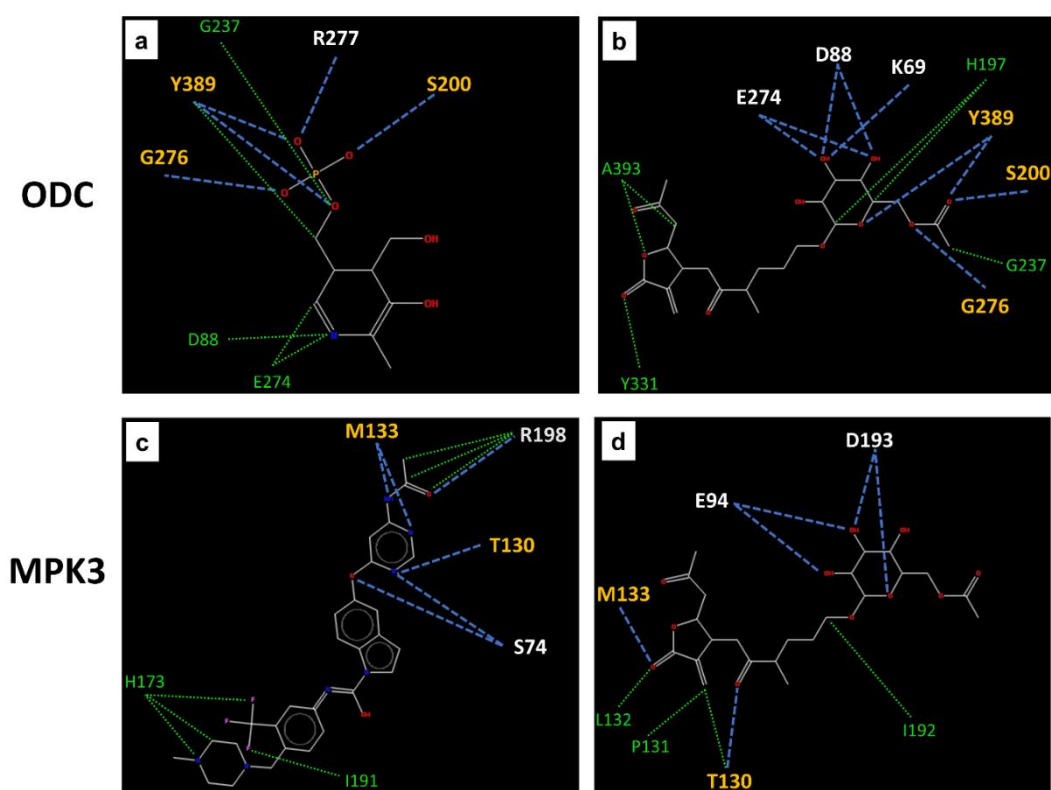


Figure 5. Hydrogen-bonding interactions (blue dotted lines) and steric interactions (green dotted lines) between (a) pyridoxal-5'-phosphate (PDB ID: PLP) and (b) disecoeudesmanolide 3A (structure **14**) at the active site of *L. donovani* ODC protein. (c) An inhibitor reported in the PDB for MPK3 (PDB ID:046) and (d) disecoeudesmanolide 3A (structure **14**), in the active site of MPK3 protein. Common H-bond interactions are highlighted in yellow.

Hydrogen-bonding (H-bond) interactions between disecoeudesmanolide 3A (structure **14**) and ODC residues are shown in Figure 5b. Six residues of the ODC active site interact with this SL: K69, D88, E274, S200, G276, and Y389. S200, G276, and Y389 also interacted with the inhibitor reported for the PDB crystal structure, pyridoxal-5'-phosphate (PDB ID: PLP, Figure 5a). In addition to the H-bond interactions between the disecoeudesmanolide 3A ester group and residues S200, G276, and Y389 of ODC, an interaction between the oxygen of the pyranose ring and Y389 was also observed. Six steric interactions, which adversely contributed to the docking energy values, were identified for structure **14**, with H197, Y331, A393, and G237; a similar steric interaction also was identified with PLP.

Similarly, structure **14** also interacts with MPK3 primarily through H-bonds and steric interactions with specific residues of the active site. Four H-bond interactions formed between E94 and D193, primarily with hydroxyl groups of the monosaccharide and the oxygen in the pyranose ring. Interestingly, two critical residues, M133 and T130, were found to form H-bonds with disecoeudesmanolide 3A (structure **14**) and with the inhibitor reported in the PDB for this protein (PDB ID: 046, Figure 5c). The interaction between the inhibitor compound and these two MPK3 residues occurs through carbonyl groups. Unlike the interactions observed for ODC, the carbonyl group of the disecoeudesmanolide 3A lactone ring interacts with MPK3, specifically at residue M133. Four steric interactions that unfavorably influence the molecular binding energy were identified for structure **14**, with the non-polar amino acids, P131, L132, and I192, which interact with the α -methylene- γ -lactone system of the SL; however, the docking score for this molecule is primarily associated with H-bonds that form with the glycoside moiety.

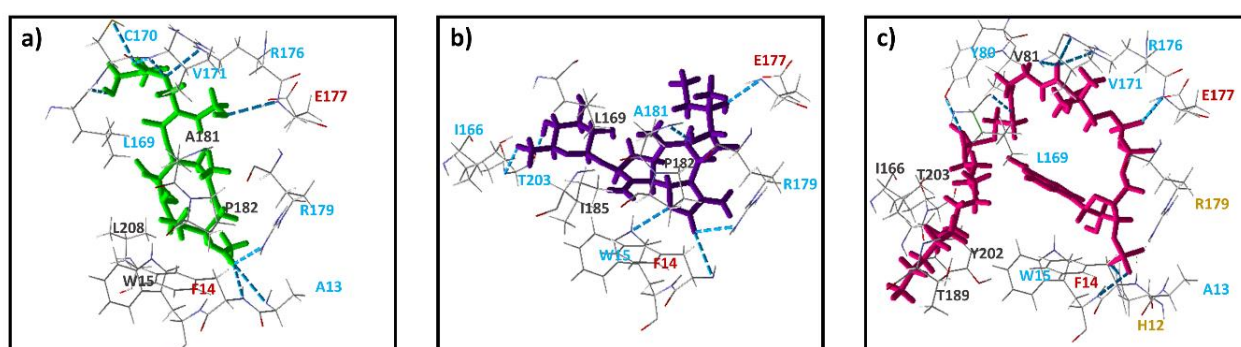


Figure 6. Docking formations between (a) 5-sarracinate-eucannabinolide (structure **11**), (b) pannonicoside (structure **12**) and (c) 2-oxopentadecyl-CoA (PDB ID: NHW) at the active site of *L. donovani* NMT. Labels: H-bonds (blue), steric interactions (gray), electrostatic interactions (yellow), and critical interactions (red).

The two best-ranked molecules for NMT were identified as 5-sarracinate-eucannabinolide (structure **11**, Figure 4) and pannonicoside (structure **12**). Figure 6 shows the primary interactions for these two SLs and NMT. Structures **11** and **12** contain a similar number of H-bonds, which may be associated with the obtained docking energy values of −98.8 and −98.1 kJ/mol, respectively. Similarly, some NMT residues were highlighted in these interactions, including F14, E177, and R179 (also establishes an electrostatic interaction with NMT), which were identified in interactions with both structures **11** and **12** and with the PDB inhibitor, 2-oxopentadecyl-CoA (PDB ID: NHW). For the two SLs, some steric interactions were also identified which may have unfavorable effects on the docking score, especially P182, which interacts with both structures. For NHW, a higher number of steric interactions were identified in the docking site for the aliphatic moiety (oxopentadecyl group). This type of interaction was not identified for the most polar region of NHW, which interacts with the *L. donovani* NMT pocket, where electrostatic interactions occur with H12 and R179.

3.3 Combined approach Structure-Based VS and Ligand-Based VS

Consensus analysis of the two methodologies used in this study (structure- and ligand-based VS) was performed, to verify potentially active molecules and their possible mechanisms of action, facilitating the identification of potential multitarget compounds. A new score (**CA**, Equation 3) was calculated, to combine the probability scores of both VS approaches and the true-negative rate of the RF model and minimize the probability of selecting false-positive compounds (Equation 3).

$$CA = \frac{SB + (1+TN) \times LB}{2+TN} \quad (\text{Equation 3})$$

where **CA** is the combined probability; **SB** is the structure-based probability; TN is the true-negative rate, and **LB** is the ligand-based probability

Different weights were assigned to the structure-based VS (**SB**, weight = 1) and ligand-based VS (**LB**, weight = 1+TN) probabilities because the structure-based probabilities were based only on interactions between proteins and ligands, whereas the ligand-based VS used pIC₅₀ experimental values and molecular descriptors to generate the RF model. Therefore, to diminish the false-positive rate (an increment of the TN), Equation 3 associated the TN rate from the internal cross-validation set with the **LB** values obtained by the SLs in each model. Minimizing the probability of selecting inactive

molecules as active molecules (false-positive) is very important for these types of studies because the selection of false-positive molecules can result in significant costs in the forms of wasted time and money [61].

Some structures that did not present with the highest scores for either of the two VS approaches appeared in the consensus analysis, as potential leishmanicidal drug structures (structures **18-23**, Figure 7). Disecoeudesmanolide 3A (structure **14**) emerged as a potential multitarget structure that may act against two target enzymes in *L. donovani* (Table 3) and was the only SL in the entire data set to demonstrate high **CA** scores for both parasitic forms against a specific protein, ODC (third-best-ranked compound for the amastigote parasitic form).

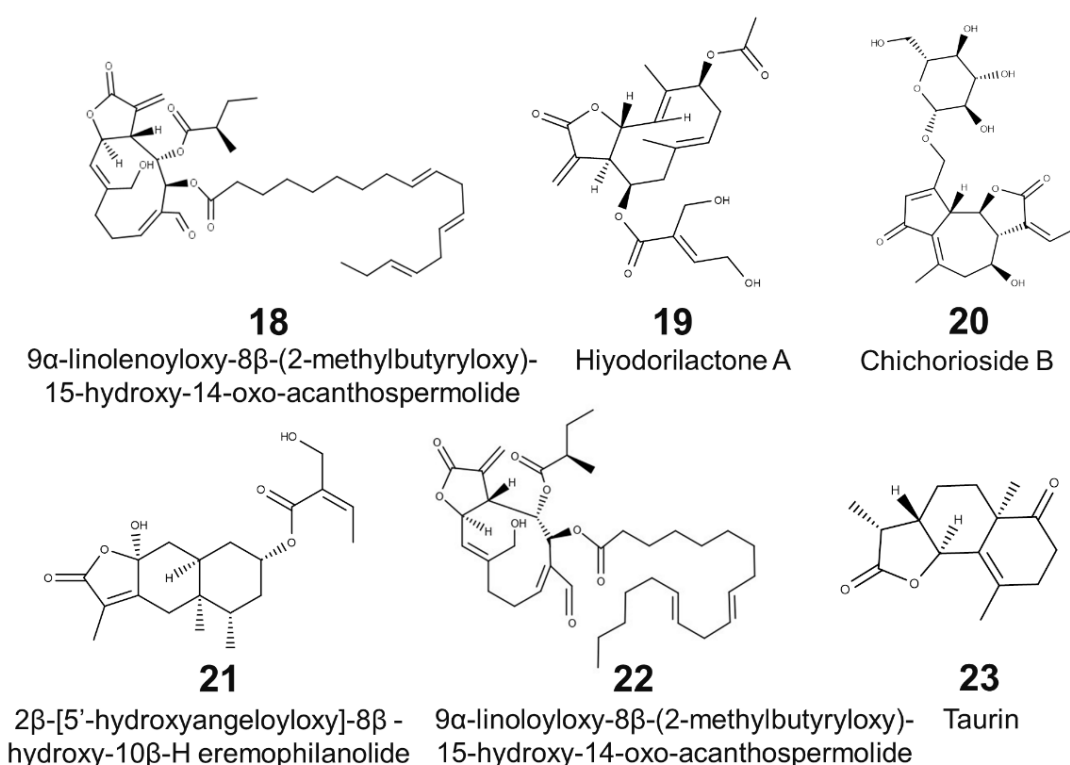


Figure 7. Representation of the best-ranked structures for each parasitic form, identified using a combined ligand-based and structure-based virtual screening approach.

Table 3. Summary of the best-ranked structures, identified using a combined ligand- and structure-based virtual screening approach; **LB** = active probability value in ligand-based VS; **SB** = active probability value in structure-based VS. **CA** = combined probability value

Protein	Amastigote				Promastigote			
	Structure	LB	SB	CA	Structure	LB	SB	CA
NMT	18	0.71	0.79	0.74	22	0.50	0.79	0.60
	19	0.75	0.68	0.73	10	0.58	0.61	0.59
ODC	13	0.61	1.00	0.75	14	0.43*	0.91	0.60
	20	0.61	0.74	0.66	-	-	-	-
MPK3	14	0.51	1.00	0.82	23	0.53	0.58	0.55
	15	0.54	0.89	0.76	-	-	-	-
LdPTR1	21	0.65	0.96	0.76	7	0.64	0.66	0.65
	2	0.78	0.67	0.74	22	0.50	0.56	0.52

Additionally, 9 α -linoloyloxy-8 β -(2-methylbutyryloxy)-15-hydroxy-14-oxo-acanthospermolide (structure 22, Figure 7), similar to SL structure 14, presented high combined probability values for two targets in *L. donovani* promastigotes (NMT and LdPTR1) and was classified as active using both VS methodologies, in this study. Structure 22, together with structure 18, which were identified as potentially active compounds against NMT in amastigotes, are both characteristic secondary metabolites of *Acanthospermum hispidum*, a plant native to Central and South America [62], and both contain a germacranolide skeleton, bound to linoleic and linolenic fatty acid ester, respectively.

For both parasitic forms, the best-ranked structures for LdPTR1, which achieved CA values above 0.70, also previously presented high probability values during the ligand-based VS: (cis, cis)-3 α -acetoxy-8 β -acetoxy-costunolide (structure 2), for amastigotes, and Bedfordia symmetric dimeric lactone (structure 7), for promastigotes. Nine compounds were classified as active using this combined approach for the promastigote parasitic form; however, only four of these SLs were also classified as active using the either of the two VS methodologies because the RF models were highly restrictive.

MD studies were performed to validate the constructed homology model and to evaluate the protein-ligand stabilities of five of the best-ranked structures identified for LdPTR1, using the various approaches in this study. Because PTR1 catalyzes the NADPH-dependent reversible reduction of oxidized pterins into dihydrobiopterin (DHB) and the

reduction of tetrahydrobiopterin (THB) and folates into 7,8-dihydrofolate (DHF) and 5,6,7,8-tetrahydrofolate (THF), the molecule 7, 8-dihydrobiopterin was used as a reference ligand for *LdPTR1* [63].

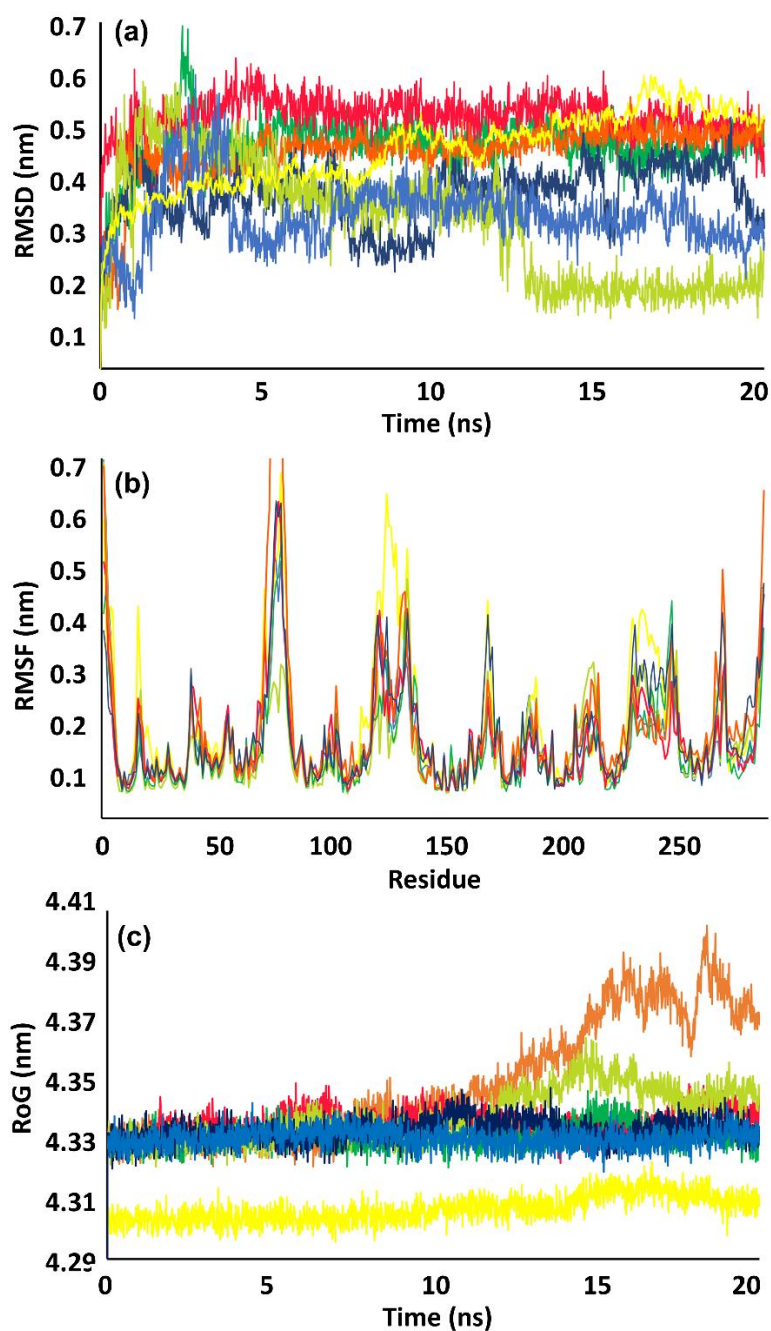


Figure 8. (a) Root mean square differentiation (RMSD), (b) root mean square fluctuation (RMSF), and (c) radius of gyration (RoG) values within the *LdPTR1* binding site, obtained after molecular dynamics simulations using five of the best-ranked SLs, identified using various VS approaches. *LdPTR1* (yellow); DHB: *LdPTR1* complex (Sky blue); structure 2: *LdPTR1* (light green); structure 7: *LdPTR1* (dark green); structure 17: *LdPTR1* (red); structure 21: *LdPTR1* (Orange); structure 22: *LdPTR1* (dark blue).

RMSD values were used to evaluate the structural stability of the receptor frame, by measuring the distance between different positions (in nm) that a set of atoms exhibited over time [64]. During the first 5 ns of the simulation, similar behaviors were observed among the 5 SLs, and *LdPTR1*, without the ligand (the apoenzyme), was less perturbed over time (Figure 8a) than the enzyme in complex with any of the other six test molecules. Starting at 7 ns, structures **2** and **22** demonstrated increased stability compared with the other SLs and the DHB ligand, whereas the *LdPTR1* apoenzyme evolved continuously and did not reach a clear stability timeframe during the 20-ns simulation.

Structure **2**, (*cis, cis*)-3 α -acetoxy-8 β -acetoxy-costunolide, which was the best-ranked molecule in the consensus analysis between the two VS methodologies, exerted more stability on the complex with *LdPTR1*, and the RMSD value for this complex was almost constant (approximately 0.15), starting at 12 ns. For the structures **7**, **17**, and **21**, the respective ligand–enzyme complexes showed similar conformational change evolutions, resulting in RMSD values of approximately 0.45, which are comparable to that obtained for *LdPTR1* alone (approximately 0.5).

Then, the RMSF values were analyzed, by examining fluctuations in each *LdPTR1* residue. The most variable regions were observed for the *LdPTR1* loops. For the section formed between A64 and T84, the ligand-*LdPTR1* complexes presented lower values than that for the apoenzyme (Figure 8b). Interestingly, structure **2** promoted reduced flexibility, in both the A64 and T84 loop region (RMSF approximately 0.35 vs. 0.75 for unbound *LdPTR1*) and the loop formed from N109 to R133 (RMSF approximately 0.7 vs. 0.3 for unbound *LdPTR1* RMSF ca), which were lower RMSF values to those for the other four SLs and DHB.

These results indicated that this compound has a better affinity for *LdPTR1* than the other molecules. Similarly, slight differences in the RMSF values were observed for the *LdPTR1* active site, except for the residues K16, S112, K198, L226, and S227, which demonstrated decreased RMSF values for the five tested SLs. Interestingly, structure **2** showed the lowest RMSF values for the whole active site, for with either *LdPTR1* or DHB: *LdPTR1*, further supporting that this SL has the highest affinity with *LdPTR1* among all of the SLs examined.

Among the SL- *LdPTR1* complexes examined (Figure 8c), fluctuations in the tertiary structure of the *LdPTR1* protein were observed (the RoG of *LdPTR1* was approximately 4.31 nm), and the complexes between *LdPTR1* and the structures **7**, **17**, and **22** appeared to be stably folded after the MD simulation (RoG of approximately 4.33 nm).

These three structures showed no differences in RoG values compared with the complex formed between DHB and *LdP*TR1. For structure **2**, a slightly major effect on *LdP*TR1 folding was observed, during the interval from 13 to 16 ns (RoG of approximately 4.36 nm); however, this fluctuation did not represent a protein deformation because the RoG values after this time (16 ns) remained similar to those for DHB and the other SLs (approximately 4.33 nm), except structure **21**, for which the MD simulation generated an unfolded protein starting at 12 ns, representing the loss *LdP*TR1 stability (RoG of approximately 4.40 nm).

Based on these MD simulations, structure **2** appears to have a better affinity for *LdP*TR1 than the other studied structures. Structural stability may be associated with the germacranolide skeleton and the presence of the α -methylene- γ -lactone moiety because the RMSD values for structures **2** and **22** (which is also a germacranolide with the same α , β -unsaturated carbonyl system) were the lowest among all examined molecules. A similar trend was observed in the structure-based VS, in which the highest **SB** value was achieved by structure **16**, which has a similar skeleton, suggesting that this type of SL should be examined in future studies attempting to identify potential antileishmanial leads.

Table 4. Binding free energies (kJ/mol) from the MM/PBSA calculations for five of the best-ranked structures identified for *LdP*TR1.

Energies		2	7	17	21	22	Ligand
Polar	Electrostatic	-18.3	-14.4	-80.1	-46.5	-27.8	-90.9
	Polar solvation	80.1	78.1	158.2	145.6	167.0	171.1
Unpolar	Van der Waals	-209.9	-235.2	-272.2	-236.5	-338.5	-113.8
	SASA	-16.6	-17.9	-22.2	-19.0	-27.8	-13.1
Binding energy		-164.7	-189.4	-216.3	-156.4	-227.1	-46.7

The binding free energies of the corresponding components can be observed in Table 4. The MM/PBSA method was used to calculate the binding free energies of the five structures that were previously analyzed by MD simulations [53]. All complexes formed with the examined SLs presented negative binding free energies, with lower values than that for the DHB-*LdP*TR1 complex. For all studied compounds, electrostatic and van der Waals interactions and the solvent-accessible surface area (SASA) contributed negatively to the binding free energies, and only polar solvation showed a positive contribution to the total energy value. DHB demonstrated a different behavior compared with the examined

SLs, with the increased contribution of polar parameters (polar solvation and electrostatic interactions) to the total binding free energy value.

In complexes with SLs, van der Waals interactions showed the highest energetic contributions, which were much stronger for structure **22** than those observed for the other four SLs, which is directly related to the presence of a very hydrophobic group, such as linoleic acid. This germacranolide presented the lowest binding energy value (–227.1 kJ/mol), confirming that it represents a promising multitarget hit against *L. donovani*. Interestingly, structure **17**, obtained a similar binding free energy value as that for structure **22**, showing that the energy calculations performed by this method are not influenced by the molecular sizes of the analyzed molecules, which is a common problem for docking calculations.

4. Conclusions

In this study, two different VS approaches were performed to identify structures with promising antileishmanial activity, from a dataset of 1,306 SLs, obtained from SistemX. For the ligand-based VS, two RF models, with accuracies above 71%, were constructed for both *L. donovani* parasitic forms. Some structural features were identified, including germacranolide and guaianolide skeletons, as the most common features among the active structures for amastigotes; in contrast, for promastigotes, dimeric molecules had the highest p-values. In the structure-based VS, using three PDB crystal structures and a homology model of *L. donovani* PTR1, SLs with lower docking scores than the reported inhibitors were identified. Finally, through a consensus analysis, the probability scores of the two VS approaches were normalized to identify 13 promising leishmanicidal SLs that were classified as being active by both VS approaches. Discoeuodesmanolide 3A (structure **14**) and 9 α -linoleoyloxy-15-hydroxy-8 β -(2-methylbutyryloxy)-14-oxo-acanthospermolide (structure **22**), appear to be promising, multitarget, antileishmanial SLs. MD simulations also showed that germacranolides with α -methylene- γ -lactone moieties, such as structures **2** and **22**, have better affinities with *LdPTR1* than other structures.

Therefore, the combined use of two VS approaches was able to preliminarily identify potentially active SLs against the two parasitic forms of *L. donovani*, from an in-house dataset. The combination of probability scores generated by the ligand- and structure-based VS approaches represents a novel methodology for these types of studies, facilitating the identification of promising molecules and their possible mechanisms of action.

References

- [1].Babuadze, G.; Farlow, J.; de Koning, H.P.; Carrillo, E.; Chakhunashvili, G.; Murskvaladze, M.; Kekelidze, M.; Karseladze, I.; Kokaia, N.; Kalandadze, I.; et al. Seroepidemiology and molecular diversity of *Leishmania donovani* complex in Georgia. *Parasit Vectors* **2016**, 9, 279, doi:10.1186/s13071-016-1558-6.
- [2].Wijerathna, T.; Gunathunga, S.; Gunathilaka, N. Recent developments and future directions in the paratransgenesis based control of *Leishmania* transmission. *Biological Control* **2020**, 104260, doi:10.1016/j.biocontrol.2020.104260.
- [3].Basano, S.d.A.; Camargo, L.M.A. Leishmaniose tegumentar americana: histórico, epidemiologia e perspectivas de controle. *Rev bras epidemiol* **2004**, 328-337, doi:10.1590/S1415-790X2004000300010
- [4].Fiocruz. As Leishmanioses. Available online: http://www.dbbm.fiocruz.br/tropical/leishman/leishext/html/ciclo_biol_gico.htm. (accessed on 29 february 2020).
- [5].Ponte-Sucre, A.; Gamarro, F.; Dujardin, J.C.; Barrett, M.P.; Lopez-Velez, R.; Garcia-Hernandez, R.; Pountain, A.W.; Mwenechanya, R.; Papadopolou, B. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. *PLoS Negl Trop Dis* **2017**, 11, e0006052, doi:10.1371/journal.pntd.0006052.
- [6].Kumar, A. Survival Strategies of *Leishmania* Parasite: Too Many Questions and Few Answers. *Current Pharmacology Reports* **2020**, 1-3, doi:10.1007/s40495-020-00209-6.
- [7].Bhattacharya, P.; Dey, R.; Dagur, P.K.; Joshi, A.B.; Ismail, N.; Gannavaram, S.; Debrabant, A.; Akue, A.D.; KuKuruga, M.A.; Selvapandian, A.; et al. Live Attenuated *Leishmania donovani* Centrin Knock Out Parasites Generate Non-inferior Protective Immune Response in Aged Mice against Visceral Leishmaniasis. *PLoS Negl Trop Dis* **2016**, 10, e0004963, doi:10.1371/journal.pntd.0004963.
- [8].WHO. World Health Organization, Leishmaniasis: Epidemiological Situation. Available online: <https://www.who.int/leishmaniasis/burden/en/> (accessed on 09 march 2020).
- [9].Anversa, L.; Tiburcio, M.G.S.; Richini-Pereira, V.B.; Ramirez, L.E. Human leishmaniasis in Brazil: A general review. *Rev Assoc Med Bras (1992)* **2018**, 64, 281-289, doi:10.1590/1806-9282.64.03.281.
- [10].Alvar, J.; Velez, I.D.; Bern, C.; Herrero, M.; Desjeux, P.; Cano, J.; Jannin, J.; den Boer, M.; Team, W.H.O.L.C. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **2012**, 7, e35671, doi:10.1371/journal.pone.0035671.

- [11].Hoet, S.; Opperdoes, F.; Brun, R.; Quetin-Leclercq, J. Natural products active against African trypanosomes: a step towards new drugs. *Nat Prod Rep* **2004**, *21*, 353-364, doi:10.1039/b311021b.
- [12].Utzinger, J.; Becker, S.L.; Knopp, S.; Blum, J.; Neumayr, A.L.; Keiser, J.; Hatz, C.F. Neglected tropical diseases: diagnosis, clinical management, treatment and control. *Swiss Med Wkly* **2012**, *142*, w13727, doi:10.4414/smw.2012.13727.
- [13].Gilbert, I.H. Drug discovery for neglected diseases: molecular target-based and phenotypic approaches: miniperspectives series on phenotypic screening for antiinfective targets. *Journal of medicinal chemistry* **2013**, *56*, 7719-7726, doi:10.1021/jm400362b.
- [14].Palma, L.C.; Ferreira, L.; Petersen, A.; Dias, B.R.S.; Menezes, J.P.B.; Moreira, D.R.M.; Hernandez, M.Z.; Veras, P.S.T. A docking-based structural analysis of geldanamycin-derived inhibitor binding to human or Leishmania Hsp90. *Sci Rep* **2019**, *9*, 14756, doi:10.1038/s41598-019-51239-0.
- [15].Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **2020**, *83*, 770-803, doi:10.1021/acs.jnatprod.9b01285.
- [16].Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M.; Scotti, M. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. *Molecules* **2017**, *22*, 79, doi:10.3390/molecules22010079.
- [17].Schmidt, T.J.; Nour, A.M.; Khalid, S.A.; Kaiser, M.; Brun, R. Quantitative structure–antiprotozoal activity relationships of sesquiterpene lactones. *Molecules* **2009**, *14*, 2062-2076, doi:10.3390/molecules14062062.
- [18].Karioti, A.; Skaltsa, H.; Kaiser, M.; Tasdemir, D. Trypanocidal, leishmanicidal and cytotoxic effects of anthecotulide-type linear sesquiterpene lactones from *Anthemis auriculata*. *Phytomedicine* **2009**, *16*, 783-787, doi:10.1016/j.phymed.2008.12.008.
- [19].Kumar, P.; Kumar, A.; Verma, S.S.; Dwivedi, N.; Singh, N.; Siddiqi, M.I.; Tripathi, R.P.; Dube, A.; Singh, N. Leishmania donovani pteridine reductase 1: biochemical properties and structure-modeling studies. *Exp Parasitol* **2008**, *120*, 73-79, doi:10.1016/j.exppara.2008.05.005.
- [20].Saudagar, P.; Saha, P.; Saikia, A.K.; Dubey, V.K. Molecular mechanism underlying antileishmanial effect of oxabicyclo[3.3.1]nonanones: inhibition of key redox enzymes of the pathogen. *Eur J Pharm Biopharm* **2013**, *85*, 569-577, doi:10.1016/j.ejpb.2013.08.014.

- [21].Shukla, A.K.; Patra, S.; Dubey, V.K. Iridoid glucosides from *Nyctanthes arbortristis* result in increased reactive oxygen species and cellular redox homeostasis imbalance in *Leishmania* parasite. *Eur J Med Chem* **2012**, *54*, 49-58, doi:10.1016/j.ejmech.2012.04.034.
- [22].Shukla, A.K.; Patra, S.; Dubey, V.K. Evaluation of selected antitumor agents as subversive substrate and potential inhibitor of trypanothione reductase: an alternative approach for chemotherapy of Leishmaniasis. *Molecular and cellular biochemistry* **2011**, *352*, 261-270, doi:10.1007/s11010-011-0762-0.
- [23].Saudagar, P.; Dubey, V.K. Cloning, expression, characterization and inhibition studies on trypanothione synthetase, a drug target enzyme, from *Leishmania donovani*. *Biol Chem* **2011**, *392*, 1113-1122, doi:10.1515/BC.2011.222.
- [24].Resh, M.D. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1999**, *1451*, 1-16, doi:10.1016/S0167-4889(99)00075-0.
- [25].Tate, E.W.; Bell, A.S.; Rackham, M.D.; Wright, M.H. N-Myristoyltransferase as a potential drug target in malaria and leishmaniasis. *Parasitology* **2014**, *141*, 37-49, doi:10.1017/S0031182013000450.
- [26].Chawla, B.; Madhubala, R. Drug targets in *Leishmania*. *J Parasit Dis* **2010**, *34*, 1-13, doi:10.1007/s12639-010-0006-3.
- [27].Nichol, C.A.; Smith, G.K.; Duch, D.S. Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. *Annu Rev Biochem* **1985**, *54*, 729-764, doi:10.1146/annurev.bi.54.070185.003501.
- [28].Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **1990**, *215*, 403-410, doi:10.1016/S0022-2836(05)80360-2.
- [29].Noble, W.S. What is a support vector machine? *Nature biotechnology* **2006**, *24*, 1565-1567, doi:10.1038/nbt1206-1565.
- [30].Fourches, D.; Muratov, E.; Tropsha, A. Curation of chemogenomics data. *Nat Chem Biol* **2015**, *11*, 535, doi:10.1038/nchembio.1881.
- [31].Muratov, E.N.; Varlamova, E.V.; Artemenko, A.G.; Polishchuk, P.G.; Kuz'min, V.E. Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol Inform* **2012**, *31*, 202-221, doi:10.1002/minf.201100129.
- [32].Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R. QSAR modeling: where have you been? Where are you going to? *Journal of medicinal chemistry* **2014**, *57*, 4977-5010, doi:10.1021/jm4004285.

- [33].Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *Journal of Molecular Structure: THEOCHEM* **2000**, *503*, 17-30, doi:10.1016/S0166-1280(99)00360-7.
- [34].Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur J Pharm Sci* **2000**, *11 Suppl 2*, S29-39, doi:10.1016/s0928-0987(00)00162-7.
- [35].Berthold, M.R.; Cebon, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter* **2009**, *11*, 26-31, doi:10.1145/1656274.1656280.
- [36].Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S.Y.; Mumper, R.J.; Tropsha, A. Quantitative nanostructure– activity relationship modeling. *ACS nano* **2010**, *4*, 5703-5712.
- [37].Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29-36, doi:10.1148/radiology.143.1.7063747.
- [38].Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* **1975**, *405*, 442-451, doi:10.1016/0005-2795(75)90109-9.
- [39].Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of medicinal chemistry* **2010**, *53*, 2719-2740, doi:10.1021/jm901137j.
- [40].Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic acids research* **2012**, *41*, D36-D42, doi:10.1093/nar/gkq1079.
- [41].Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys* **1978**, *185*, 584-591, doi:10.1016/0003-9861(78)90204-7.
- [42].Gourley, D.G.; Schuttelkopf, A.W.; Leonard, G.A.; Luba, J.; Hardy, L.W.; Beverley, S.M.; Hunter, W.N. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. *Nat Struct Biol* **2001**, *8*, 521-525, doi:10.1038/88584.
- [43].Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **1993**, *234*, 779-815, doi:10.1006/jmbi.1993.1626.

- [44].Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **1993**, *26*, 283-291, doi:10.1107/S0021889892009944.
- [45].Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Ca geometry: ϕ , ψ and $\text{C}\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics* **2003**, *50*, 437-450, doi:10.1002/prot.10286.
- [46].Diego, S. Discovery Studio Modeling Environment Release 2.1. *Accelrys Software Inc* **2007**.
- [47].Dufe, V.T.; Ingner, D.; Heby, O.; Khomutov, A.R.; Persson, L.; Al-Karadaghi, S. A structural insight into the inhibition of human and *Leishmania donovani* ornithine decarboxylases by 1-amino-oxy-3-aminopropane. *Biochemical Journal* **2007**, *405*, 261-268, doi:10.1042/BJ20070188.
- [48].Brannigan, J.A.; Smith, B.A.; Yu, Z.; Brzozowski, A.M.; Hodgkinson, M.R.; Maroof, A.; Price, H.P.; Meier, F.; Leatherbarrow, R.J.; Tate, E.W.; et al. N-myristoyltransferase from *Leishmania donovani*: structural and functional characterisation of a potential drug target for visceral leishmaniasis. *J Mol Biol* **2010**, *396*, 985-999, doi:10.1016/j.jmb.2009.12.032.
- [49].Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; 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.
- [50].Berendsen, H.J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer physics communications* **1995**, *91*, 43-56, doi:10.1016/0010-4655(95)00042-E.
- [51].Schüttelkopf AW, V.A.D. PRODRG: a tool for high-throughput crystallography of proteinligand complexes. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, doi:10.1107/S0907444904011679.
- [52].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. *Journal of computational chemistry* **2004**, *25*, 1605-1612, doi:10.1002/jcc.20084.
- [53].Kumari, R.; Kumar, R.; Open Source Drug Discovery, C.; Lynn, A. g_mmpbsa • A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling* **2014**, *54*, 1951-1962.

- [54].Castillo-González, D.; Mergny, J.-L.; De Rache, A.; Pérez-Machado, G.; Cabrera-Pérez, M.A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G.F.; Guédin, A.; Bourdoncle, A. Harmonization of QSAR best practices and molecular docking provides an efficient virtual screening tool for discovering new G-quadruplex ligands. *Journal of chemical information and modeling* **2015**, *55*, 2094-2110, doi:10.1021/acs.jcim.5b00415.
- [55].Rueda, M.; Abagyan, R. Best Practices in Docking and Activity Prediction. *bioRxiv* **2016**, 039446, doi:10.1101/039446.
- [56].Scotti, M.T.; Herrera-Acevedo, C.; Oliveira, T.B.; Costa, R.P.O.; Santos, S.; Rodrigues, R.P.; Scotti, L.; Da-Costa, F.B. Sistemax, an Online Web-Based Cheminformatics Tool for Data Management of Secondary Metabolites. *Molecules* **2018**, *23*, 103, doi:10.3390/molecules23010103.
- [57].Geissman, T.; Griffin, S.; Waddell, T.G.; Chen, H.H. Sesquiterpene lactones. Some new constituents of *Ambrosia* species: *A. psilostachya* and *A. acanthicarpa*. *Phytochemistry* **1969**, *8*, 145-150, doi:10.1016/S0031-9422(00)85806-9.
- [58].Schmidt, T.J. Toxic activities of sesquiterpene lactones: structural and biochemical aspects. *Curr. Org. Chem* **1999**, *3*, 4.
- [59].Popławski, J.; Holub, M.; Samek, Z.; Herout, V. On terpenes. CCIX. Arnicolides-sesquiterpenic lactones from the leaves of *Arnica montana* L. *Collection of Czechoslovak Chemical Communications* **1971**, *36*, 2189-2199, doi:10.1135/cccc19712189.
- [60].Xie, C.; Wang, H.; Sun, X.; Meng, L.; Wang, M.; Bartlam, M.; Guo, Y. Isolation, Characterization, and Antiproliferative Activities of Eudesmanolide Derivatives from the Flowers of *Inula japonica*. *J Agric Food Chem* **2015**, *63*, 9006-9011, doi:10.1021/acs.jafc.5b03075.
- [61].Acevedo, C.H.; Scotti, L.; Scotti, M.T. In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018**, *13*, 634-645, doi:10.1002/cmdc.201700743.
- [62].Walker, R.H.; Wells, L.W.; McGUIRE, J.A. Bristly starbur (*Acanthospermum hispidum*) interference in peanuts (*Arachis hypogaea*). *Weed Science* **1989**, *37*, 196-200, doi:10.1017/S0043174500071770.
- [63].Kaur, J.; Dube, D.; Ramachandran, R.; Singh, P.; Singh, N. Thianthrene is a novel inhibitor of *Leishmania donovani* pteridine reductase 1 (PTR1). *Journal of molecular biochemistry* **2012**, *1*.
- [64].Rozo-Lugo, C.; Cuca-Suárez, L.E.; Schmidt, T.J.; Coy-Barrera, E. Tetrahydrobenzofuran-6 (2 H)-one Neolignans from *Ocotea heterochroma*: Their Platelet

Activating Factor (PAF) Antagonistic Activity and in Silico Insights into the PAF Receptor Binding Mode. *Journal of natural products* **2018**, *81*, 1968-1975, doi:10.1021/acs.jnatprod.8b00189.



CAPÍTULO IV

A diidrofolato redutase-timidilato sintase (DHFR-TS) é uma enzima crucial no metabolismo do folato, dependendo do NADPH para sua função. Desempenha um papel fundamental na geração de tetraidrofolato a partir do folato ou diidrofolato e facilita a conversão de monofosfato de deoxiuridina (dUMP) em monofosfato de deoxitimidina (dTMP) [1]. Simultaneamente, a enzima PTR1, com uma função semelhante à DHFR, desempenha um papel chave na regulação do equilíbrio redox e na homeostase dos níveis de folato, contribuindo para contornar a inibição metabólica da DHFR convertendo diidrofolato (DHF) em tetraidrofolato (THF). Esse mecanismo leva à ineficácia de medicamentos contra *Leishmania* e ao desenvolvimento de resistência medicamentosa no parasita [2].

Foi relatado que a inibição da diidrofolato redutase-timidilato sintase (DHFR-TS) faz com que a Pteridina redutase 1 (PTR1) forneça folato suficiente para garantir a sobrevivência do parasita. Portanto, ambas as enzimas, DHFR-TS e PTR1, devem ser consideradas como potenciais alvos terapêuticos na busca por tratamentos mais eficazes contra a leishmaniose [3]. No entanto, essa teoria ainda é debatida, uma vez que a eliminação do gene *ptr1* provou ser letal para o parasita, indicando o papel crucial de PTR1 no crescimento e na metaciclogênese do parasita devido à produção reduzida de pterina [4].

Este capítulo busca a seleção de potenciais inibidores de PTR1 e DHFR-TS em diferentes espécies de *Leishmania* causadoras de Leishmaniose cutânea (CL), a partir de um banco de moléculas in-house composto por 360 cauranos. Em todo o mundo, aproximadamente, 20 diferentes espécies de *Leishmania* são responsáveis pela transmissão de CL, incluindo *L. tropica*, *L. major*, *L. aethiopica*, *L. infantum* e *L. donovani*, sendo que *L. major* representa o organismo causador mais comum [5]. No Novo Mundo (do sul dos Estados Unidos à América Latina e América do Sul), *L. mexicana*, *L. venezuelensis*, *L. amazonensis*, *L. braziliensis*, *L. panamensis*, *L. guyanensis* e *L. peruviana* são as principais espécies causadoras de CL [6,7].

O capítulo está dividido em duas partes. Na primeira, é criado um modelo classificatório de random forest para a seleção de potenciais inibidores contra *L. major* PTR1. O modelo foi validado por meio de ensaios *in vitro* usando a enzima recombinante *L. major* PTR1. Adicionalmente, considerando que a maioria dos casos de CL na América Latina é causada por *L. amazonensis*, *L. braziliensis* e *L. panamensis*, foram realizados cálculos computacionais, incluindo docking molecular e simulações de dinâmica

molecular, para avaliar a capacidade desses cauranos de inibir em diferentes espécies do parasita.

A segunda parte do capítulo buscou avaliar se os compostos selecionados contra PTR1 também podem ter uma capacidade multialvo, sendo também inibidores de DHFR-TS. Foram realizados ensaios *in vitro* dos cauranos previamente selecionados, e foi encontrada uma potencial capacidade inibitória dupla DHFR-TS/PTR1 para os compostos: ácido 3-*p*-cumaroiloxi-ent-caur-16-eno-19-oico e ácido 3-cinamoiloxi-ent-caur-16-eno-19-oico.

REFERÊNCIAS

- [1]. BAFANDEH, Elnaz et al. Expression analysis of DHFR and PTR1 genes in *Leishmania major* expose to olive leaf extract. **South African Journal of Botany**, v. 157, p. 520-524, 2023.
- [2]. TEIXEIRA, Bárbara Velame Ferreira et al. Dual and selective inhibitors of pteridine reductase 1 (PTR1) and dihydrofolate reductase-thymidylate synthase (DHFR-TS) from *Leishmania chagasi*. **Journal of Enzyme Inhibition and Medicinal Chemistry**, v. 34, n. 1, p. 1439-1450, 2019.
- [3]. VICKERS, Tim J.; BEVERLEY, Stephen M. Folate metabolic pathways in *Leishmania*. **Essays in biochemistry**, v. 51, p. 63-80, 2011.
- [4]. NARE, Bakela et al. New approaches to *Leishmania* chemotherapy: pteridine reductase 1 (PTR1) as a target and modulator of antifolate sensitivity. **Parasitology**, v. 114, n. 7, p. 101-110, 1997.
- [5]. GHATEE, Mohammad Amin; TAYLOR, Walter R.; KARAMIAN, Mehdi. The geographical distribution of cutaneous leishmaniasis causative agents in Iran and its neighboring countries, a review. **Frontiers in public health**, v. 8, p. 11, 2020.
- [6]. PAZ, Carlos et al. *Leishmania major*, the predominant *Leishmania* species responsible for cutaneous leishmaniasis in Mali. **The American journal of tropical medicine and hygiene**, v. 88, n. 3, p. 583, 2013.
- [7]. VOJTKOVA, Barbora et al. Central Asian rodents as model animals for *Leishmania major* and *Leishmania donovani* research. **Microorganisms**, v. 8, n. 9, p. 1440, 2020.

The background of the page is a complex, abstract pattern. It features a network of interconnected nodes and lines, resembling a molecular structure or a data network. The nodes are represented by small circles in various shades of blue and grey, while the lines are thin and light grey. Interspersed among these network elements are several hexagonal shapes, some of which are solid blue and others are outlined in blue. The overall composition is dense and layered, with elements appearing to recede into the background, creating a sense of depth. The color palette is primarily monochromatic, using different tones of blue and grey on a white background.

Parte A

Identification of Kaurane-Type Diterpenes as Inhibitors of *Leishmania* Pteridine Reductase I

Chonny Herrera-Acevedo ^{1,4}, Areli Flores-Gaspar ^{2,*}, Luciana Scotti ¹, Francisco Jaime Bezerra Mendonça-Junior ³, Marcus Tullius Scotti ^{1,*} and Ericsson Coy-Barrera ^{2,4}

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa, PB 58051-900, Brazil; chonny622@gmail.com; luciana.scotti@gmail.com; mtscotti@gmail.com.

² Departamento de Química, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia; areli.flores@unimilitar.edu.co.

³ Laboratory of Synthesis and Drug Delivery, State University of Paraíba, João Pessoa, Brazil; franciscojbmendonca@yahoo.com.br.

⁴ Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia; ericsson.coy@unimilitar.edu.co.

Artigo publicado em *Molecules* (IF: 4.927)

DOI Number: <https://doi.org/10.3390/molecules26113076>

Ano de publicação: 2021

Abstract: The current treatments against *Leishmania* parasites present high toxicity and multiple side effects, which makes the control and elimination of leishmaniasis challenging. Natural products constitute an interesting and diverse chemical space for the identification of new antileishmanial drugs. To identify new drug options, an *in-house* database of 360 kauranes (tetracyclic diterpenes) was generated, and a combined ligand- and structure-based virtual screening (VS) approach was performed to select potential inhibitors of *Leishmania major* (*Lm*) pteridine reductase I (PTR1). The best-ranked kauranes were employed to verify the validity of the VS approach through *Lm*PTR1 enzyme inhibition assay. The half-maximal inhibitory concentration (IC₅₀) values of selected bioactive compounds were examined using the random forest (RF) model (i.e., 2β-hydroxy-menth-6-en-5β-yl *ent*-kaurenoate (**135**) and 3α-cinnamoyloxy-*ent*-kaur-16-en-19-oic acid (**302**)) were below 10 μM. A compound similar to **302**, 3α-*p*-coumaroyloxy-*ent*-kaur-16-en-19-oic acid (**302a**), was also synthesized and showed the highest activity against *Lm*PTR1. Finally, molecular docking calculations and molecular dynamics

simulations were performed for the VS-selected, most-active kauranes within the active sites of PTR1 hybrid models, generated from three *Leishmania* species that are known to cause cutaneous leishmaniasis in the new world (i.e., *L. braziliensis*, *L. panamensis*, and *L. amazonensis*) to explore the targeting potential of these kauranes to other species-dependent variants of this enzyme.

Keywords: *Leishmania*; Natural products; Kauranes; Asteraceae; Virtual screening; Machine learning; Molecular docking

1. Introduction

Leishmaniasis refers to a group of anthroponotic and zoonotic diseases that affect between 700,000 and 1 million people worldwide, causing between 20,000 and 30,000 deaths each year, primarily among populations found in tropical and subtropical areas. Leishmaniasis has been classified as a neglected tropical disease (NTD) due to a lack of research and the poor development of new drugs over many decades. [1-3]. Leishmaniasis is caused by approximately 20 protozoan parasite species of the genus *Leishmania*, which are transmitted to humans by more than 30 different species of phlebotomine sandflies [4]. The distinct species of *Leishmania* cause at least four separate syndromes: visceral leishmaniasis (VL, also known as kala-azar), post-kala-azar dermal leishmaniasis (PKDL), cutaneous leishmaniasis (CL), and mucocutaneous leishmaniasis (MCL) [5].

The CL subtype is typically characterized by localized, diffuse, or disseminated skin lesion[6]. In the old world (southern Europe, the Middle East, southwest Asia, and Africa), approximately 20 different *Leishmania* species are responsible for the transmission of CL, including *L. tropica*, *L. major*, *L. aethiopica*, *L. infantum*, and *L. donovani*, with *L. major* representing the most common causative organism [7]. In the new world (from the southern United States through Latin America to South America), *L. mexicana*, *L. venezuelensis*, *L. amazonensis*, *L. braziliensis*, *L. panamensis*, *L. guyanensis*, and *L. peruviana* are the primary causal species of CL [8, 9]. In Colombia, the overall leishmaniasis incident rate is 26.2 cases per 100,000 population (including 98.6% of the cases related to CL), and in Brazil, autochthonous cases of CL have been reported in all states. Colombia and Brazil represent the new world countries with the most frequently reported CL clinical manifestation [10, 11].

Starting in the 1950s, pentavalent antimonial compounds were introduced as treatments against *Leishmania* species; however, these drugs are associated with several

adverse events and are becoming increasingly ineffective due to the development of resistance [12, 13].

Other drugs used to treat leishmaniasis include amphotericin B in a liposomal formulation, which significantly reduced the side effects and treatment duration associated with amphotericin B in the free form but is very expensive; and paromomycin and miltefosine, which are associated with high toxicity (particularly renal toxicity), increased resistance, and teratogenic and abortifacient effects [4, 12]. Therefore, alternative chemotherapies must be developed to improve the control and elimination of this group of diseases. Natural products, which have always been an important source of bioactive compounds, are commonly used as the starting material for new drug development [14-16].

Recently, computational studies using natural products have been reported in the continuous search for new leishmanicidal drugs or lead compounds. In particular, machine learning and molecular docking calculations have been used to identify new structures with potential anti-*Leishmania* activities, based on secondary metabolites found in Asteraceae species [17, 18], especially sesquiterpenoids [19, 20], triterpenes [21] and phytosterols [22]. However, anti-*Leishmania* studies examining the effects of diterpenoids, a common class of secondary metabolites found in Asteraceae (more than 1,200 structures have been identified), are rare [23].

Thus, in this study, an *in silico* approach, combining both structure- and ligand-based virtual screening (VS), was used to select structures with potential activity against pteridine reductase 1 (PTR1) from *L. major* (LmPTR1) from an in-house database containing 360 kauranes. PTR1 (E.C. 1.5.1.33), is an NADPH-dependent short-chain reductase, is responsible for the unusual salvage of pterin in *Leishmania* and acts as a metabolic bypass for drugs that target dihydrofolate reductase [24].

Subsequently, the *in silico* results were verified through *in vitro* assays, determining the half-maximal inhibitory concentrations (IC₅₀) for the structures **135**, **301**, and **302**. In addition, two derivatives structures (**301a** and **302a**) were synthesized, and their IC₅₀ values were also calculated. Finally, molecular docking and molecular dynamics simulations were performed to identify potential kauranes against PTR1 of various *Leishmania* species known to cause CL in the new world.

2. Results and discussion

2.1 A combined ligand-/structure-based virtual screening approach using LmPTR1.

2.1.1. Ligand-based VS

The ChEMBL dataset (<https://www.ebi.ac.uk/chembl/>) was classified as either active or inactive (binary classification), using a cutoff value of $\text{pIC}_{50} \geq 5.0$ ($\text{pIC}_{50} = \log \text{IC}_{50}$). This value was selected according to the range of pIC_{50} values observed for the entire dataset (657 structures) to obtain the maximum representation of the chemical space for each class of structure (active and inactive). Structures with pIC_{50} values between 4.9 and 5.0 (range of 0.1 units) were excluded to avoid edge effects and improve the predictive capacity of the models by minimizing potential activity differences associated with errors and different experimental protocols. IC_{50} values describe the concentration of a given substance required to inhibit 50% of parasite growth [20].

Subsequently, VolSurf+ descriptors (128) were calculated for the remaining molecules, including 298 inactive (46.9%) and 338 active (53.1%) molecules. All VolSurf+ descriptors [25, 26] together with their respective binary classifications were used to build a random forest (RF) model in Knime software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org) [27]. A model with 200 trees was selected, and the Gini Index was used as a split criterion, which has the lowest false-positive rate. A five-fold cross-validation procedure was performed, splitting the dataset five times into a modeling set (80%/20%). Only the modeling set, which was additionally divided into multiple training and test sets (80%/20%), was used to build and validate the models [28].

For the training set used in the RF model, the match percentage values approached 100%. Sensitivity (true-positive rate) values of 78.1% and 82.6 % and specificity (true-negative rate) values of 72.7% and 73.7%, were obtained for the cross-validation and test sets, respectively. Two parameters were calculated to evaluate the quality of the RF model: the receiver operating characteristic (ROC) curve and Matthews's correlation coefficient (MCC). The area under the ROC curve (AUC) plots the true-positive rate (sensitivity) against the false-positive rate ($1 - \text{specificity}$), and the MCC correlates all values in the confusion matrix [29, 30].

For the *L. major* RF model, AUC values of 0.85 and 0.87 were obtained for the internal cross-validation and five-fold external validation datasets, respectively. When

calculating the MCC parameter, a value of 1 represents a perfect prediction, a value of 0 represents a random prediction, and a value of -1 represents total disagreement between the prediction and the observation. Our *L. major* RF model returned values of 0.51 (cross-validation) and 0.57 (five-fold external validation) [30]. The slightly higher MCC value obtained for the five-fold external validation (0.57) demonstrates a high degree of differentiation between the active and inactive compounds identified in the ChEMBL dataset.

The applicability domain (APD) was used to assess the reliability of the predictions for the samples in the test and SL sets, and the calculation of the APD is based on the molecular interactions determined by the VolSurf+ descriptors [14, 20]. For the *L. major* RF model set, more than 98.4% of molecules were classified as reliable, with only 8 molecules classified as unreliable. When the RF models were applied to the kaurane dataset, more than 94.2% of molecules were classified as reliable in each model, with only 20 molecules classified as unreliable. Unreliable molecules were removed.

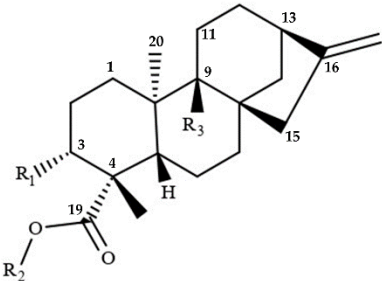
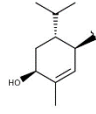
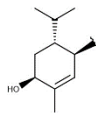
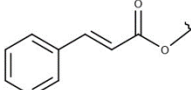
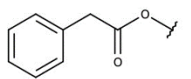
	ID	R ₁	R ₂	R ₃	LB
	135	H		H	0.57
	134	H		$\Delta^{9,11}$	0.55
	302		H	H	0.54
	298		H	H	0.53

Figure 1: Potentially active kauranes, identified using RF model (ligand-based VS), for *L. major*. LB active probability value.

A ligand-based VS was performed on the remaining 340 kauranes was performed. Only 7 of the 340 structures were classified as active (ligand-based probability value [LB] ≥ 0.5), with structures **134** and **135** representing the two best-ranked kauranes, with LB values of 0.57 and 0.55, respectively (Figure 1). These two diterpenoids are found in *Wedelia chinensis*, a species of Asteraceae [31]. Structurally, these two kauranes are characterized by the presence of (1S,4R,5R)-2-Methyl-5-propan-2-ylcyclohex-2-ene-1,4-

diol, linked through an ester bound to the kaurenoic acid. The *LB* values for these two kauranes are almost identical, indicating that the activity of these two compounds is likely associated with the presence of this monoterpenoid and the pi-bond in the structure between C9 and C11.

Additionally, two additional kauranes isolated from *Wedelia trilobata*, structures **298** and **302**, also presented *LB* values greater than 0.5. Cinnamoyl (**302**) and 2-phenylacetic (**298**) esters are established with the carboxyl group of the kaurenoic acid (Figure 1) [32]. In these two structures, the functional groups present in R3 were also found to play key roles, as structure **301**, which also includes a cinnamoyl ester, was classified as inactive (*LB* = 0.48). The presence of a hydroxyl moiety in R3 represents the unique structural difference between structures **301** and the active structure **302**.

2.1.2. Structure-based VS

A structure-based VS (molecular docking) was performed to explore the mechanism of action of the kauranes dataset against the crystal structure of PTR1 (E.C. 1.5.1.33), an NADPH-dependent short-chain reductase that is responsible for the unusual salvage of pterin in *Leishmania* and acts as a metabolic bypass for drugs that target dihydrofolate reductase [24]. The docking scores and the respective deviation values for the best-ranked structures are reported in Table 1 (all binding energy values can be found in Supplementary Material, Table S3). All tested molecules were ranked using the following probability calculation (Equation 1), as previously reported by Herrera-Acevedo et al. [14, 20]. Those kauranes that presented structure-based probability values (*SB*) above 0.5 were classified as active.

$$SB = (E_i/E_{min}) > 0.5 \text{ and } E_i < E_{ligand} \quad \text{Equation 1}$$

where ***SB*** is the structure-based probability; E_i is the docking energy of compound *i*, where *i* ranges from 1 to 360 (Kauranes dataset); E_{min} is the lowest energy value of the dataset; and E_{ligand} is the ligand energy from protein crystallography.

The docking results showed that all 360 compounds obtained ***SB*** values above 0.5; however, relative to the PTR1 inhibitors that were used as controls, 252 structures and 359 structures had lower docking scores than 7,8-dihydro-L-biopterin (DHB) and pyrimethamine (PMA), respectively. The Protein Data Bank (PDB) ligand for *Lm*PTR1, methotrexate (PDB ID: MTX) [33], has a calculated docking score of −560.4 kJ/mol..

Table 1: Docking energies of the best-ranked structures from the structure-based VS for *L. major* PTR1. SD = Standard Deviation; RMSD values = Root Mean Square Deviation and **SB** = Structure-based probability.

Ligand	Docking Score (kJ/mol)	SD	RMSD	SB
101	-449.5	2.8	1.5	1.00
270	-437.6	7.4	1.6	0.97
302	-423.0	9.4	1.3	0.94
299	-422.7	9.2	1.3	0.94
175	-421.8	18.0	1.0	0.94
298	-420.2	20.1	1.6	0.93
174	-419.9	9.7	1.4	0.93
173	-419.7	7.4	1.3	0.93
135	-416.7	9.1	1.1	0.93
MTX	-560.4	17.6	0.4	-

Structures **135** and **302** (Figure 1), which were predicted to have high *LB* probability values based on the RF model, also showed high *SB* values and were two of the ten best-ranked kauranes identified, with *SB* values of -423.0 kJ/mol and -416.7 kJ/mol, respectively. Spatially, in the active site of *LmPTR1*, structures **135** and **302** adopted an *L-shaped* conformation, similar to that observed for the ligand methotrexate (Figure 2a). Base on the two-dimensional analysis, common interactions were identified for these two kauranes compared with methotrexate, highlighting the π -alkyl interaction with M233 and the van der Waals interactions with S112, Y191, K198, and G225 (Figure 2).

Methotrexate achieved a docking score of -560.4 kJ/mol in the active site of *LmPTR1*, and the formation of two H-bond interactions with S111 and N118 were observed (Figure 2b). Structure **302** also formed two H-bonds between S227 and the carboxylic group of C-19. Additionally, the aromatic ring of F113 interacted with both **135** and **302**, in addition to methotrexate, through π - π and π -alkyl interactions. Two steric interactions that unfavorably influenced the molecular binding energy were identified for the structures **135** (R17 and D232) and **302** (S111 and L226), as shown in Figure 2c and d, respectively.

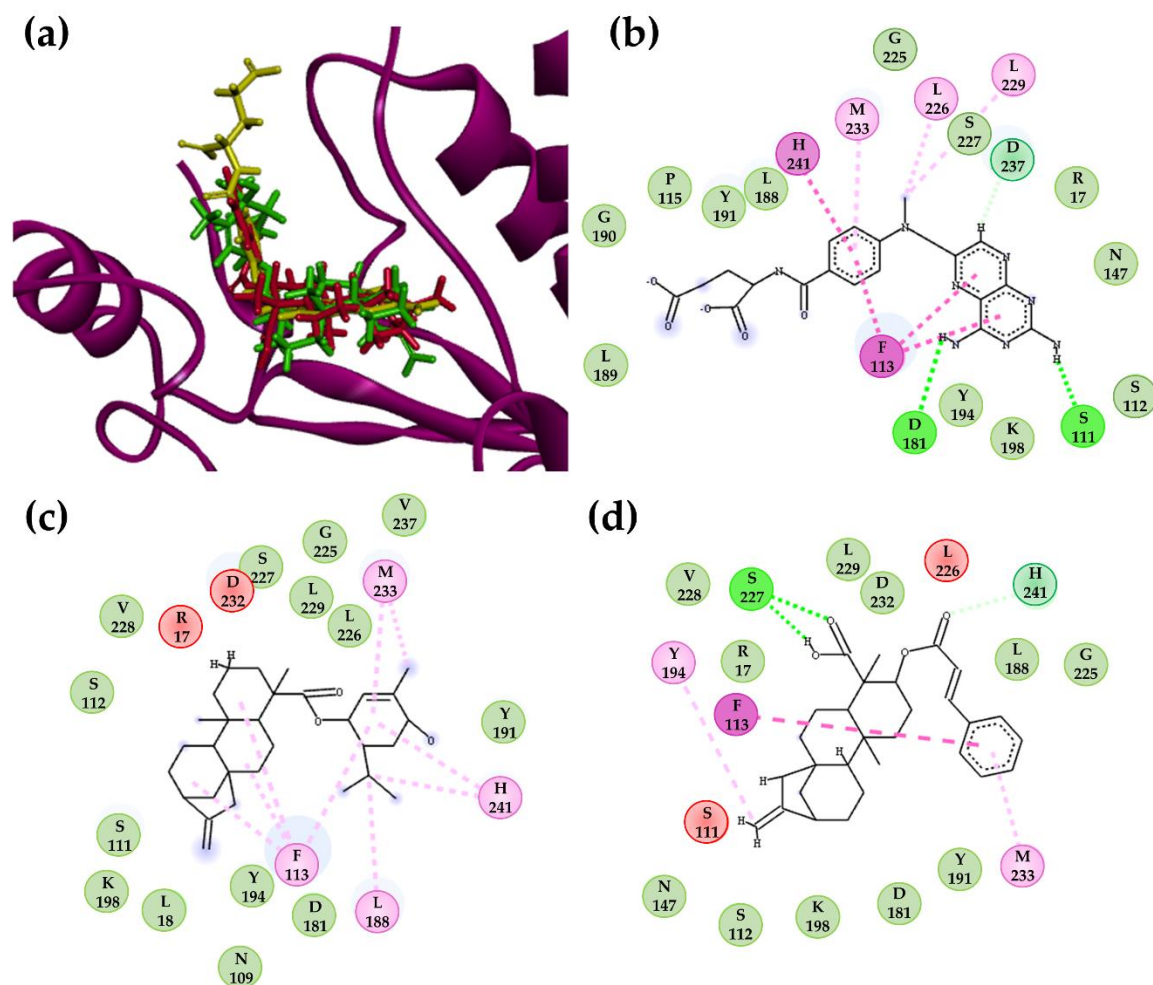


Figure 2: a) Docking conformations of Structure **135** (green), **302** (red) and MTX (yellow) in the active site of *LmPTR1*; 2D-residual interaction diagrams of (b) methotrexate (MTX), (c) structure **135** and (d) structure **302**. Interacting residues are shown as colored circles depending on the interactions (as colored dashed lines): H-bond (lime), Van der Waals (green), π - π (purple) and π -alkyl (pink), unfavorable (red) and Carbon H-bond (teal) interactions.

2.1.3. Consensus analysis of the two VS approaches.

To verify the potentially active kauranes and their possible mechanisms of action, a combined approach using both structure- and ligand-based VS was performed. An equation was used to combine the probability scores of both VS approaches with the true-negative rate from the RF model to minimize the probability of selecting false-positive compounds (Equation 2) [14, 20].

$$CA_{Lm} = [SB + (1+TN) \times LB] / [2 + TN]$$

Equation 2

where CA_{Lm} = combined-approach probability, SB = structure-based probability, TN = true-negative rate, and LB = ligand-based probability.

Equation 2 is based on the fact that the ligand-based VS uses pIC_{50} experimental values; thus, the LB score has a high weight with respect to the SB score, which only relates interactions between the protein and ligand. The ligand-based VS seeks to reduce the probability of selecting inactive molecules as active compounds (false positive); therefore, in Equation 2, the LB is associated with an increment of TN .

Table 2: Kauranes classified as active using an approach combining ligand-based and structure-based VS.

Kaurane	SB	LB	CA_{Lm}
135	0.93	0.57	0.70
101	1.00	0.51	0.69
302	0.94	0.54	0.68
134	0.90	0.55	0.68
298	0.93	0.53	0.68

Table 2 shows the results for the five kauranes that were classified as active using the combined approach and the two VS methodologies. Four of the five structures that previously displayed a high active probability value in the ligand-based VS (Figure 1) emerged as interesting potential anti-*Leishmania* structures that might act on the *LmPTR1* protein.

In addition, fischericin F (structure **101**), extracted from *Ligularia fischeri*, a species of the *Ligularia* genus (Asteraceae) [34], was also classified as potentially active in the combined approach ($CA_{Lm} = 0.69$). Although this kaurane did not present the highest scores from the ligand-based VS, it emerged as the best-ranked structure from the structure-based VS approach. Structurally, **101** has ferulic acid as the main feature, bound to the ent-kaurane skeleton through an ester bond at C14.

Through this combined approach, based on two different VS methodologies, five kauranes from various Asteraceae species were identified as having promising antileishmanial activity against *LmPTR1* from a dataset of 360 kauranes, with structures **302** and **135** indicated as having high probability values based on both the ligand-based and structure-based VS approaches.

2.2. *In vitro* enzymatic activity inhibition for VS-selected kauranes against LmPTR1.

To verify the results obtained from the combined approach using the two VS methodologies, the *in vitro* enzymatic inhibitory activities of structures **135** and **302** (actives) and structure **301** (inactive) were examined. In addition, two kauranes, structures **301a** and **302a** (in which the cinnamoyloxy group was replaced by a coumaroyloxy group), were also tested against LmPTR1. The diterpenes **135**, **301**, and **302** were synthesized for use in an *in vitro* enzymatic activity inhibition assay. This aim was oriented to identify appropriate precursors, as such compounds are not commercially available. Therefore, a phytochemical study was initially performed focusing on the fruits of *Xylopia frutescens*, an annonaceous plant that is rich in kaurane-type diterpenes [35].

This procedure led to the isolation of various diterpenes, but interest was focused on *ent*-kaurenoic acid (**A**), 3 α -hydroxy-*ent*-kaur-16-en-19-oic acid (**B**), and 3 α ,9 β -dihydroxy-*ent*-kaur-16-en-19-oic acid (**C**) because these structures are fully elucidated by spectroscopic data interpretation, and comparisons are possible with data available in the literature [36, 37]. Therefore, these compounds were considered suitable precursors to initiate the synthesis of target compounds. Thus, compound **135** was obtained from the commercially available (*R*)-(-)-carvone (**D**) (Figure 3), which was first transformed into 5 β -hydroxy-(*R*)-carvone (**E**) by chemoselective monohydroxylation and subsequently reduced to 2-oxo-menth-6-en-5 β -ol (**F**) by selective hydrogenation using the Wilkinson's catalyst [38].

Diterpenic acid **A** esterified with **F** under mild conditions via Steglich esterification [39] to produce 2-oxo-menth-6-en-5 β -yl *ent*-kaurenoate (**G**). This monoterpene/diterpene ester adduct was finally converted into 2 β -hydroxy-menth-6-en-5 β -yl *ent*-kaurenoate (**135**) through the selective 1,2 reduction of α,β -unsaturated ketones using Luche conditions [40], in which the *Re* face of the enone in **G** favored the desired β -epimer (68% epimeric excess).

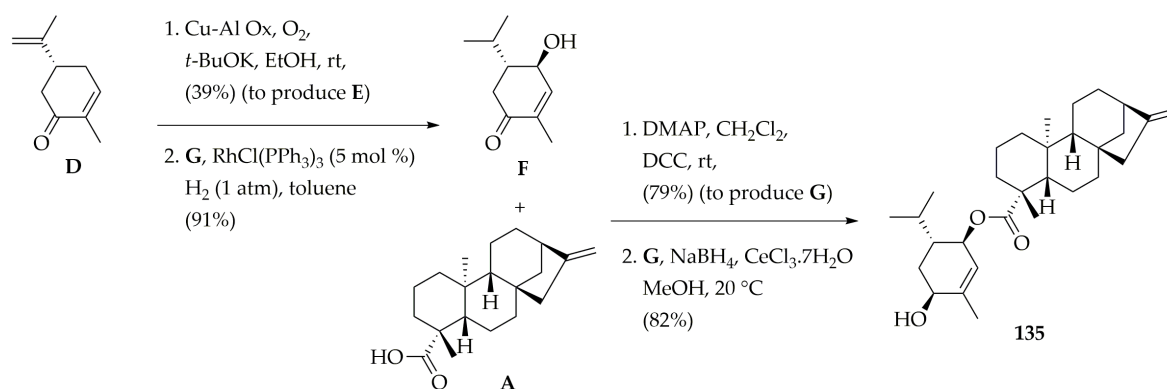


Figure 3. Synthetic route to produce monoterpene/diterpene ester adduct **135**.

Steglich esterification was also exploited to obtain the other two selected diterpenes (Figure 4). Isolated compounds **B** and **C** were separately esterified with cinnamic acid (**H**), yielding the phenylpropanoid/diterpene ester adducts **301** and **302**, respectively, with good yields (78%–79%). Additionally, the scope of this reaction was expanded to produce compounds **301a** and **302a**, using the same diterpene precursors (**B** and **C**) condensed with *p*-coumaric acid (**I**) to observe the influence of the *p*-hydroxyl group at the phenylpropanoid moiety in the subsequent enzymatic study.

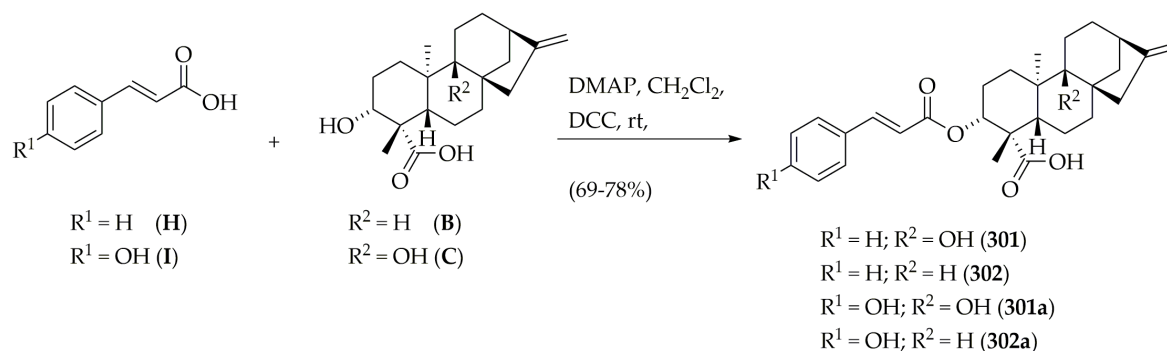


Figure 4. Synthetic route to produce phenylpropanoid/diterpene ester adducts **301**, **302**, **301a**, and **302a**.

The selected synthetic diterpene esters **135**, **301**, **302**, **301a**, and **302a** were tested *in vitro* to experimentally determine their abilities to inhibit the enzymatic activity of *LmPTR1* as an extension of the results provided by the *in silico* screening. Recombinant *LmPTR1* was kinetically assessed, as previously reported [40], to ensure the appropriate enzymatic features, resulting in a consistent substrate *K_m* of 5.6 μM. After testing *LmPTR1* inhibition, the selected diterpenes exhibited inhibitory properties at different levels, following a concentration-response behavior within the 0.1–128 μM range. The IC₅₀ was then calculated for the tested diterpenes, and these values were used to calculate the apparent inhibitory constant (*K_i^{app}*) (Table 3) using the Cheng-Prusoff equation, assuming reversible competitive inhibition and 1:1 stoichiometry [41]. PMA, a known *PTR1* inhibitor, was used as a positive control.

Among the three VS-selected diterpenes, **135** was found to be the most potent inhibitor, whereas **301** exhibited the lowest *K_i^{app}*. Remarkably, the inhibitory activity was improved by approximately 60% if a 3α-*p*-coumaroyloxy group was present in **302** instead of a 3α-cinnamoyloxy substituent, as **302a** exhibited a lower *K_i^{app}* value than **302**. No similar effect was observed for **301**, as **301a** showed a slightly lower inhibitory activity

than **301**. Therefore, a reasonable inference based on this small set of compounds is that the presence of a *p*-hydroxyl group at the phenylpropanoid moiety might favor inhibitory activity, whereas a 9 β -hydroxyl group at the diterpene moiety has a negative influence on *Lm*PTR1 inhibition.

Finally, although the test diterpenes were found to be less active than the positive control, the concentration-response behavior and the consequently calculated K_i^{app} (≤ 5 μ M) of the selected diterpenes demonstrated the validity of the designed VS approach for the selection of bioactive compounds against PTR1 and the computationally-studied binding modes of these selected compounds within the active site of *Lm*PTR1, which is associated with the development of CL. These selected compounds can be considered important leads that can be used to obtain additional active PTR1 inhibitors.

Table 3: Results for VS-selected diterpenes as inhibitors of *Lm*PTR1

Compound	135	302	301	302a	301a	PMA
IC ₅₀ (μ M)	8.6	9.6	21.2	6.1	23.2	1.11
Confidence Interval (95%)	9.4–7.9	10.7–8.6	23.4–18.9	7.1–5.2	26.3–20.4	1.20–1.01
K_i^{app}	1.88	2.10	4.64	1.33	5.08	0.24

2.3. Molecular docking calculations for the kaurane dataset using hybrid models of *La*, *Lb*, and *Lp*PTR1

The structures **135**, **302**, and **302a** displayed *in vitro* activity against *L. major*, which is one of the species responsible for most CL cases in the Mediterranean littoral, the Middle East, the Indian subcontinent, and central Asia [42]. However, in the American continent, other *Leishmania* species, such as *L. amazonensis* (*La*), *L. braziliensis* (*Lb*), and *L. panamensis* (*Lp*), are associated with great clinical diversity, associated particularly with CL and MCL [43]. Therefore, the potential activity of kauranes against PTR1 from these three species must also be examined, despite the absence of crystal structure for these species.

2.3.1. Hybrid models of *La*, *Lb*, and *Lp*PTR1

Hybrid models were built in YASARA software (YASARA (18.4.24) Vienna, Austria: YASARA Biosciences GmbH; 2018) [44] from sequences of three *Leishmania* species, *Lp*, *La*, and *Lb*. To verify and validate the reliability and stereochemical qualities

of the modeled protein, data from Ramachandran, WHAT IF, and VERIFY 3D plots and the quality Z-scores of dihedrals were determined for the built models, which describes how many standard deviations separate the model quality from the average high-resolution X-ray structure [45]. Higher values are better, and negative values indicate that the homology model looks worse than a high-resolution X-ray structure [46, 47]. The Ramachandran plot showed that the main possible chain conformations included more than 88.7% of residues in the most favored regions for the three hybrid models, with close to 10.0% of residues in allowed regions. Only the *Lp* model showed two residues (0.8%) in disallowed regions (outliers; Supplementary Material, Figure S2). Because the percentage of residues found in the outlier region was low or absent, the generated models were considered satisfactory. Eleven residues in the active site were analyzed against the template sequence and were found to be conserved [33].

According to the VERIFY 3D results (<https://services.mbi.ucla.edu/SAVES/>), 87.1% (*Lp*), 86.1% (*Lb*), and 80.0% (*La*) of residues had mean 3D/1D scores ≥ 0.2 , which indicated a reliable model because more than 80% of amino acids had values of 0.2 in the 3D/1D profile. The verification of dihedral quality was classified as optimal for the three models, with values above 0.913. The quality of atomic contacts between the atoms of each residue was analyzed using the Coarse Packing Quality Control module of WHAT IF (<https://swift.cmbi.ru.nl/servers/html/index.html>), which compares the distribution of atom positions around each residue. The mean scores of all wastes were -0.334 , -0.488 , and -0.667 , for *Lb*, *La*, and *Lp*, respectively. A score of less than -5.0 for a residue indicates poor or unusual atomic contacts.

2.3.2. Molecular docking calculations for kauranes dataset.

Molecular docking calculations for the 360 kaurane dataset plus the two derivative compounds, **301a** and **302a**, were obtained using the Autodock/Vina algorithm for the three generated *Leishmania* hybrid models (*Lp*, *Lb*, and *La*) to evaluate whether the kauranes that showed *in vitro* activity against *L. major* have the potential to display multispecies activity. Equation 3 combines the *SB* probability scores obtained from the docking calculations of all three models, and DHB and PMA were used as references.

$$CA = [(LaSB + LbSB + LpSB) / 3] \geq 0.5 \quad \text{Equation 3}$$

where *LbSB* is the structure-based probability score for *L. braziliensis*, *LpSB* is the structure-based probability score for *L. panamensis*, and *LaSB* is the structure-based probability score for *L. amazonensis*. *CA* is the consensus analysis for all three species.

Table 4: Kauranes classified as active using an approach combining ligand-based and structure-based VS.

Kaurane	<i>Lb SB</i>	<i>Lp SB</i>	<i>La SB</i>	<i>CA</i>
302a	0.87	1.00	1.00	0.96
301a	0.86	0.97	0.98	0.94
175	0.90	0.95	0.92	0.92
69	0.93	0.94	0.88	0.92
135	1.00	0.88	0.82	0.90
134	0.93	0.80	0.87	0.87
302	0.85	0.89	0.82	0.86

Therefore, a *CA* value equal to or greater than 0.5 is classified as active. Among the 362 structures tested, only 274 were classified as active, and **301a** and **302a** were the best-ranked molecules, with *CA* values of 0.96 and 0.94, respectively. The kauranes (**135** and **302**) that demonstrated *in vitro* activity against *L. major* also showed high *CA* values (above 0.86) and were among the ten best-ranked molecules (Table 4).

DHB showed more affinity for PTR1 from the three *Leishmania* species than PMA. Lower docking scores than the two control structures were obtained for 100%, 81%, and 99% of the tested kauranes for *Lb*, *Lp*, and *La*, respectively.

Docking poses for structure **302** in the active site of the three *Leishmania* PTR1 models and the interacting residues for **302**, DHB, and PMA are displayed in Figure 5 and Table 5, respectively. A Vina score of -9.73 kcal/mol was calculated for *Lp*, predominantly due to van der Waals interactions, with five common interactions identified between DHB and PMA (L19, S112, Y194, L226, and S227). A critical common π -anion interaction was observed between D181 and the aromatic ring of the cinnamoyl group. No H-bond interactions are observed for this kaurane in the active site of *Lp*PTR1.

Similarly, the structure of **302** achieved a Vina score of -11.1 kcal/mol in the active site of *Lb*PTR1, exhibiting some common van der Waals interactions with DHB and PMA (S112, S227, and L228. An H-bond interaction was established between G225 and the carboxylic group of C-19. G225 did not interact with DHB and PMA, which establish three

H-bonds (L19 and N110 were common between these two molecules). Interestingly, an alkyl interaction with L19 was observed for the structures **302** and **135**, which was the best-ranked molecule for *LbPTR1* (Vina score of -13.07 kcal/mol).

For structure **302**, in the active site of *LaPTR1*, two H-bond interactions were observed with A15 and K16, and K16 was also observed in the complex between *LaPTR1* and DHB, identified as a critical amino acid for the binding. For both the kaurene **302** and the two controls (PMA and DHB), a higher number of van der Waals interactions were exhibited than any other type of intermolecular interaction, although only the interaction with Y193 was common among all three of these structures. Finally, an alkyl interaction with P223 was identified for the structures **302** and PMA.

Table 5: VINA scores and interactions of structure **302**, PMA and DHB with aminoacid residues of *LpPTR1*, *LbPTR1* and *LaPTR1*. Critical interactions are highlighted in bold font.

Protein	Ligand	VINA score (kcal/mol)	Interacting residues
<i>LpPTR1</i>	Structure 302	-9.73	<i>Van der Waals</i> : A14, G20, L19 , N110, S112 , Y114, M179, I180, Q186, P187, Y194 , G225, L226 , S227 , L228, F229, Y283; <i>Carbon H-bond</i> : K198; <i>Alkyl</i> : R18, L19; <i>π-alkyl</i> : M183; <i>π-sigma</i> : L188; <i>π-anion</i> : D181 .
	PMA	-7.92	<i>H-bond</i> N110, I180; <i>Van der Waals</i> : R18, L19 , S112 , M179, Y194 , K198, G225, L226 , S227 , L228; <i>π-alkyl</i> : Y114, F229; <i>π-π T-shaped</i> : Y114; <i>π-anion</i> : D181 .
	DHB	-8.33	<i>H-bond</i> : M179, D181, K198, G224; <i>Van der Waals</i> : L19, S112, Y194, P224, L226, S227, F229; <i>Carbon H-bond</i> : I180; <i>π-π T-shaped</i> : Y114 <i>π-anion</i> : D181 .
<i>LbPTR1</i>	Structure 302	-11.1	<i>H-bond</i> G225 ; <i>Van der Waals</i> : K17, R18, N110, S112 , Y114, I180, D181, L188, Y194, K198, S227 , L228 , F229, P230, Y241; <i>π-sigma</i> : M233, L226; <i>Alkyl</i> : L19
	PMA	-7.41	<i>H-bond</i> : R14, L19, N110; <i>Van der Waals</i> : G20, C21, A111, S112 , S227 , L228 ; <i>π-alkyl</i> : R18, Y194; <i>π-sigma</i> : Y114.
	DHB	-7.75	<i>H-bond</i> : L19, N110, P224; <i>Van der Waals</i> : A14, K17, R18, G20, C21, S112 , I179, I180, D181, A182, Y194, S227 , L228 .

LaPTR1	Structure 302	-9.87	<i>H-bond</i> : A15, K16 ; <i>Van der Waals</i> : T12, G13, A14, R17, L18, H36, Y37, H38, R39, S40, N109, S111, S146, Y193 , K197; <i>Alkyl</i> : P223 <i>π-alkyl</i> : A110, L66.
	PMA	-7.19	<i>H-bond</i> : G224; <i>Van der Waals</i> : S111, M178, V179, A181, Y193 , L228, M232; <i>π-alkyl</i> : P223 ; <i>Alkyl</i> : F113, L187, L225, Y240 <i>π-anion</i> : D180.
	DHB	-7.61	<i>H-bond</i> : K16 , R17, N109; <i>Van der Waals</i> : G13, G19, M178, V179, D180, A181, Y193 , K197, P223, G224, L225; <i>π-alkyl</i> : R17, L18.

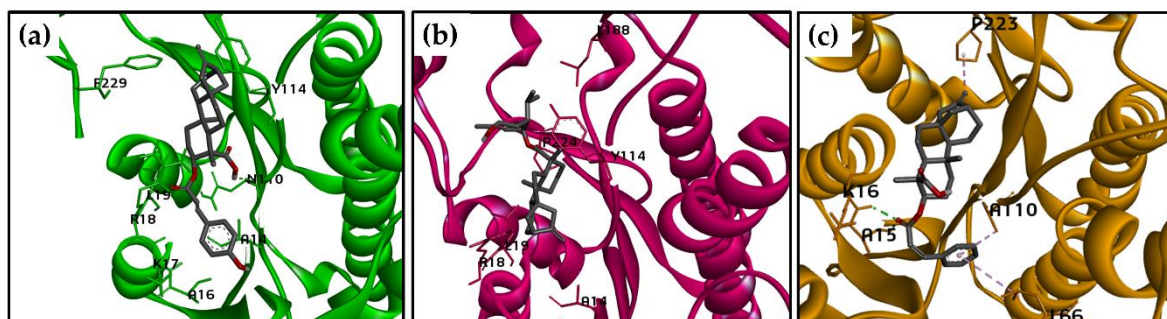


Figure 5. Docking formations between a structure **302** in the active site of a) *LpPTR1*; b) *LbPTR1* and c) *LaPTR1*. Labels: H-bonds (green), π - alkyl interactions (purple).

2.4. Molecular dynamics simulations

L. braziliensis is the causative agent of human CL and MCL in various countries of the American continent, including Colombia, Brazil, Nicaragua, and Ecuador, among others [43, 48, 49]. Thus, to validate the hybrid model constructed for *LbPTR1* and to evaluate the protein–ligand stabilities of the structures **135**, **302**, and **302a**, molecular dynamics (MD) studies were performed using DHB and PMA as reference ligands.

Root-mean-square deviation (RMSD) values were used to evaluate the structural stability of the receptor frame by measuring the distance between different positions (in nm) that a set of atoms exhibited over time [50]. In the first half of the simulation time (0–25 ns), the structures **135**, **302**, **302a**, DHB, PMA, and the apoenzyme of *LbPTR1* (apo*LbPTR1*, protein without ligand) showed a similar grade of perturbation, with RMSD values ranging from approximately 0.35 to 0.65 nm. After 25 ns, all ligands exhibited reduced perturbations relative to that observed for apo*LbPTR1*, which suggests increased stability exerted by the inhibitors on the complex with *LbPTR1*. RMSD values for the

protein–kaurane complexes of approximately 0.5 nm were observed, except for the reference ligand, DHB, which showed a slightly higher RMSD value (approximately 0.55 nm). In contrast, apo*LbPTR1* exhibited values approaching 0.7 nm (Figures 6a).

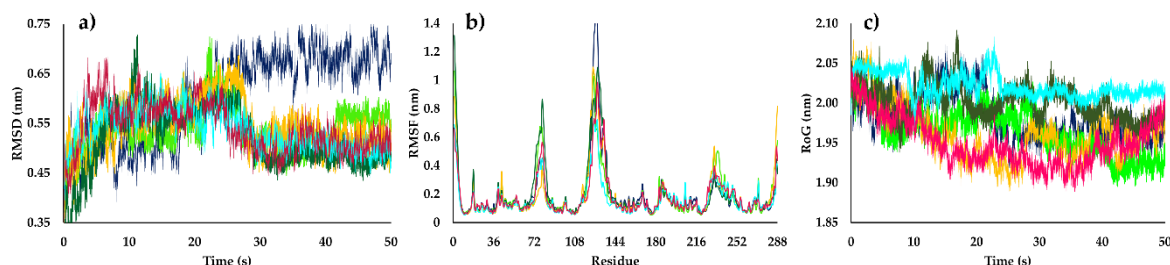


Figure 6. a) Root-mean-square deviation (RMSD), b) root-mean-square-fluctuation (RMSF), and c) radius of gyration (RoG) values within the *LbPTR1* binding site, obtained after molecular dynamics simulations using three of the best-ranked structures in molecular docking calculations. apo*LbPTR1* (blue); DHB: *LbPTR1* complex (light green); PMA: *LbPTR1* complex (sky blue); structure **135**: *LbPTR1* complex (pink); structure **302**: *LbPTR1* complex (yellow); structure **302a**: *LbPTR1* complex (dark green);

The fluctuations for each *LbPTR1* residue were analyzed by calculating root-mean-square fluctuation (RMSF) values. Kauranes, DHB, and PMA in complex with *LbPTR1* presented lower values than the apoenzyme, and the *LbPTR1* loops were identified as the most variable regions. In the sections of *LbPTR1* with defined tertiary structures (helical or β -sheets), the fluctuation of residues for both the apoenzyme and the complexes formed with DHB, PMA, and kauranes (**135**, **302**, and **302a**) was less than 0.25 nm. For most of the residues in the active site, the RMSF values decreased when *LbPTR1* was in complex with structure **302**.

In the loop region, from A65 to S85, structure **302a** showed the highest RMSF value (approximately. 0.9 nm) compared with structures **135** and **302a**, which reached RMSF values lower than 0.6 nm. This behavior might be due to differences in the spatial conformation of **302a** within the active site of *LbPTR1* compared with those for structures **135** and **302**; consequently, the molecular docking values are justified. The analysis of the loop section between N110 to T135 showed that inhibitors (structures **135**, **302**, and **302a**, DHB, and PMA) in complex with *LbPTR1* reached RMSF values approaching 1.0 nm; in contrast, the apoenzyme exhibited a value above 1.65 nm (Figure 6b), indicating that these structures stabilized the protein following the formation of an *LbPTR1*-kaurane complex.

The diterpenes showed similar RMSF values as DHB; however, in this loop region, PMA has a lower RMSF value (approximately 0.8 nm).

The evolution of the *LbPTR1* packing level was observed through the radius of gyration (RoG) values. The diterpene-*LbPTR1* complexes showed no differences in RoG values compared with the complex formed between DHB and *LbPTR1* (RoG of approximately 2.00 nm), with fluctuations in the tertiary structure lower than 0.10 nm. The RoG values for PMA were slightly different (approximately 2.05), demonstrating a different behavior throughout the 50 ns test period than the other structures analyzed (Figure 6c). Apo*LbPTR1*, during the initial 25 ns, showed a higher RoG than the complexes, with an RoG value approaching 2.05 nm. However, in the second half of the simulation, a decrease in the RoG value was observed, reaching a value similar to those observed for the complexes formed with diterpenes and DHB (RoG of approximately 1.95 nm). Based on these results, the structures **135**, **302**, and **302a** appeared to be stably folded after the MD simulation.

According to the MD simulations, the binding free energies for structures **135**, **302**, **302a**, PMA, and DHB were calculated through the MM/PBSA method. The diterpenes **135**, **302**, and **302a** presented binding free energy values of -132.7 kJ/mol, -121.4 kJ/mol, and -138.3 kJ/mol, respectively, which were all lower energy values than those measured for DHB and PMA in complex with *LbPTR1*, which presented free energy values of -107.4 kJ/mol and -110.0 kJ/mol, respectively. These differences in energetic contributions were associated with structural differences (Table 6).

Table 6: Binding free energies (kJ/mol) from the MM/PBSA calculations for three of the best-ranked structures identified for *LbPTR1*; DHB and PMA were used as reference ligands

	135		302		302a		PMA		DHB	
Energy contribution	kJ/mol	SD	kJ/mol	SD	kJ/mol	SD	kJ/mol	SD	kJ/mol	SD
Van der Waals	-210.7	6.0	-170.8	7.9	-208.6	7.6	-138.8	1.7	-121.3	3.0
Electrostatic	-2.9	1.5	-26.7	3.4	-9.7	3.0	-145.0	2.5	-194.6	10.3
Polar solvation	103.6	4.1	95.5	9.9	100.7	13.1	186.4	5.9	221.4	12.0
SASA	-22.7	0.5	-19.4	0.9	-20.6	0.4	-12.7	0.4	-12.9	0.3
Binding energy	-132.7	7.6	-121.4	6.1	-138.3	9.3	-110.0	4.2	-107.4	6.1

In the three kauranes, van der Waals interactions showed the highest negative contributions to the binding free energy, which supported the previously observed docking results. The solvent-accessible surface area (SASA) and electrostatic parameters contributed negatively, but to a lesser degree, to the binding free energies in similar proportions (except for the electrostatic parameter of **302**, which displayed a higher contribution to total binding energy).

The reference inhibitor (PMA) and the native ligand (DHB) demonstrated different behaviors from those observed for the three diterpenes, with a higher contribution of electrostatic interactions to the total binding free energy value, which represented the energy parameter with the highest negative energetic contribution. For all molecules, the polar solvation had a positive contribution to the total energy value, with larger contributions to the complexes DHB-*Lb*PTR1 and PMA-*Lb*PTR1.

3. Materials and Methods

3.1. Database

From the ChEMBL database (<https://www.ebi.ac.uk/chembl/>), we selected a diverse set of 1,085 structures that were initially classified according to their predicted activity against *L. major*. These compounds were classified according to their pIC₅₀ values [$-\log\text{IC}_{50}$ (mol/L)]; therefore, we stratified them into active (pIC₅₀ $\geq$ 5.0) and inactive (pIC₅₀ < 5.0) structures.

The APD, based on Euclidean distances, was used to identify those compounds in the test set for which predictions may be unreliable; compounds were considered unreliable if they had APD values higher than $d + Z\sigma$, where d is the average Euclidian distance, and σ is the standard deviation of the set of samples used as the training set, with lower-than-average Euclidian distance values relative to all samples in the training set. The parameter Z is an empirical cutoff value, and 0.5 was used as the default value [51].

Structures with pIC₅₀ values ranging from 4.9 to 5.0 (range of 0.1 units) were excluded to avoid edge effects and improve the predictive capacity of the models. Excluding these structures minimized the differences in activity values resulting from errors and differences in experimental protocols [52]. Data curation was performed for the datasets according to procedures suggested in the literature [53-55]. Standardizer software [JChem, version 16.11.28 (2016), calculation module developed by ChemAxon, <https://www.chemaxon.com/>] was used to canonize all simplified molecular-input line-entry system (SMILES) codes. After duplicate structures were removed, those with higher

pIC₅₀ values were eliminated, facilitating the generation of more restrictive models. Finally, 638 structures for *L. major* (338 active and 300 inactive structures) were included in the analysis,

The kaurane dataset was built in-house, and a total of 360 molecules from this dataset were used in this study. For all structures, SMILES codes were used as the input data in Marvin [ChemAxon, version 16.11.28 (2016), calculation module developed by ChemAxon, <https://www.chemaxon.com/>]. We used standardizer software [JChem, version 16.11.28 (2016), calculation module developed by ChemAxon, <https://www.chemaxon.com/>]. ChemAxon was used to canonize the structures, add hydrogens, perform aromatic form conversions, and clean molecular graphs in three dimensions.

3.2. Volsurf+ descriptors

The three-dimensional (3D) structures of the identified molecules, in special data file (SDF) format, were used as the input data for VolSurf+, v. 1.0.7 [25] and were subjected to molecular interaction fields (MIFs) to generate descriptors using the following probes: N1 (amide nitrogen–hydrogen bond donor probe), O (carbonyl oxygen–hydrogen bond acceptor probe), OH2 (water probe), and DRY (hydrophobic probe). Additional non-MIF-derived descriptors were generated, resulting in a total of 128 descriptors [25]. One of the main advantages of using VolSurf descriptors is the relatively low influence of conformational sampling and averaging on these descriptors [26].

3.3. RF models

Knime 3.1.0 software (KNIME 3.1.0 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org) [27] was used to perform all of the following analyses. Initially, the descriptors calculated in the VolSurf+ program were imported in comma-separated value (CSV) format, and the “Partitioning” node in the stratified sampling option was used to classify 80% of the initial dataset as the training set and the remaining 20% as the test set.

The model was generated by employing the modeling set and the RF algorithm, with a five-fold external validation procedure, using WEKA nodes. In the five-fold cross-validation procedure, the dataset is divided five times into a modeling set (80%/20%). The modeling set (which was used to build and validate the models) was further divided into training (80%) and test sets (20%) [28, 53]. The parameters selected for the RF models

included the following: number of trees to build = 200; seed for random number generator = 1; and Gini Index, as a split criterion, for both the training and internal cross-validation sets.

From the confusion matrix, the internal and external performances of the selected models were analyzed, using the following parameters: sensitivity (true-positive rate), specificity (true-negative rate), and accuracy (overall predictability). In addition, to describe the true performance of the model with more clarity than can be obtained from accuracy alone, the ROC curve was employed, using a “ROC curve” node, which uses the sensitivity and specificity parameters. The plotted ROC curve shows the true-positive (active) rate versus the false-positive rate ($1 - \text{specificity}$) [29].

In this representation, when a variable of interest cannot be distinguished between the two groups, the AUC value is 0.5, whereas a perfect separation between the values of the two groups, with no distribution overlap, results in an AUC value of 1. The MCC was also calculated, for which a value of 1 represents a perfect prediction, a value of 0 represents a random prediction, and a value of -1 represents a total disagreement between the prediction and the observation [30].

3.4. False-positive remover

To detect false-positive structures among the SLs that were classified as active in the two RF models, the substructure filter for the removal of pan assay interference compounds (PAINS) was used [56]. All SMILES codes for SLs classified as active were submitted to PAINS removal (<https://www.cblligand.org/PAINS/>), and those structures that were classified as false-positives were excluded from the final analysis.

3.5 Synthesis of VS-selected diterpenes

3.5.1. Materials and reagents

Optical rotations and UV data were recorded using a Jasco P-2000ST digital polarimeter and a Thermo Fisher Scientific Genesys 10S spectrophotometer, respectively. ^1H and ^{13}C Nuclear magnetic resonance experiments were recorded in a Bruker Avance400 spectrometer using CDCl_3 as solvent. All shifts are given in δ (ppm) using the signal of TMS as reference. All coupling constants (J) are given in Hz. HRESIMS data were obtained on a Bruker micro-QToF II spectrometer, respectively. Thin-layer chromatography (TLC) using silica gel 60 F254 TLC plates (Merck) and mobile phases comprising solvent mixtures of n-hexane, EtOAc, and MeOH were used. Plates after TLC

development were observed under UV light (254 and 365 nm) and derivatized using I2 vapor and Hannessian' s reagent (aqueous solution of ammonium molybdate, cerium sulphate and H2SO4). Silica gel (SiO2) 60 (0.04-0.063 mm) (Merck) was used for flash chromatography (flash CC). Cinnamic acid, p-coumaric acid, (R)-(-)-carvone and other reagents and solvents for synthesis and enzyme inhibition assay were acquired from Sigma-Aldrich.

3.5.2. Isolation of compounds A-C.

Fruits (325 g) of *Xylopia frutescens* (Annonaceae) were extracted with ethanol 96% and a portion of the resulting crude extract (25.5 g) was fractionated by CC over SiO2 in gradient elution (n-hexane to methanol) affording twenty-five different fractions. Fractions 7, 8, 11 and 13 were independently depurated by flash CC on SiO2, yielding compounds ent-kaurenoic acid (A) (75.6 mg) [37] 3 α -hydroxy-ent-kaur-16-en-19-oic acid (B) (52.3 mg) [36] and 3 α ,9 β -dihydroxy-ent-kaur-16-en-19-oic acid (C) (42.6 mg), using mixtures of n-hexane:EtOAc:MeOH 9:1.5:0.5; 8:1:1, 7:2:1 and 6:2:2, respectively.

3 α ,9 β -dihydroxy-ent-kaur-16-en-19-oic acid (C): Oil; $[\alpha]_D^{20} = -61.8$ (c = 0.1, CHCl₃); ¹H NMR (400 MHz, CHCl₃) δ 4.78 (br s, 1H), 4.67 (br s, 1H), 4.63 (dd, J = 12.0, 5.0 Hz, 1H), 2.65 (br s, 1H), 2.63 (br d, J = 13.5 Hz, 1H), 2.49-2.44 (m, 1H), 2.31 (dd, J = 11.0, 1.3 Hz, 1H), 2.16-2.10 (m, 2H), 2.07-2.03 (m, 1H), 1.88-1.85 (m, 1H), 1.82-1.76 (m, 2H), 1.73-1.67 (m, 1H), 1.65-1.55 (m, 4H), 1.48-1.44 (m, 2H), 1.35-1.33 (m, 1H), 1.29 (s, 3H), 1.15 (s, 3H); ¹³C NMR (100 MHz, CHCl₃) δ 180.1, 156.8, 104.4, 78.6, 76.9, 50.3, 49.4, 48.5, 48.5, 43.6, 42.1, 39.6, 35.2, 34.5, 30.5, 29.6, 24.3, 24.2, 20.9, 17.4; HREIMS $[M+H]^+$ m/z 335.2203 (calcd for C₂₀H₃₁O₄, 335.2222).

3.5.3. Synthesis of 2 β -hydroxy-menth-6(1)-en-5 β -yl ent-kaurenoate (135)

The synthesis of the top-ranked ester 135 was accomplished following the next four synthetic steps:

3.5.3.1. Synthesis of 5 β -hydroxy-(R)-carvone (E)

Compound G was synthesized as previously reported [38]. Briefly, Cu–Al Ox catalyst (168 mg) was placed into a 100-mL round-bottom flask (RBF) containing absolute EtOH (30 mL). The resulting mixture was stirred at room temperature (rt) for 10 min. Subsequently, (R)-(-)-carvone (D) (450 mg, 3.0 mmol, 1.0 equiv) and t-BuOK (168 mg, 1.5 mmol, 0.5 equiv) were added, and this reaction mixture was further stirred at rt for 30

h. After completion, the mixture was filtered through celite, rinsing with MeOH (15 mL). The filtrate was concentrated under reduced pressure, and the residue was purified by flash CC on SiO₂ (10% EtOAc in n-hexane) to afford compound E (194 mg, 39% yield). Oil; $[\alpha]_{\text{D}}^{20} +65.2$ (c 0.1, CHCl₃); ¹H NMR (400 MHz, CDCl₃) δ 6.85 (br d, J = 9.7 Hz, 1H), 5.05 (br s, 2H), 4.48 (dd, J = 9.7, 1.9 Hz, 1H), 3.29 (ddd, J = 6.5, 2.6, 1.9 Hz, 1H), 2.65 (dd, J = 11.8, 2.6 Hz, 1H), 2.29 (dd, J = 11.8, 6.5 Hz, 1H), 1.79 (s, 3H), 1.77 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) 198.5, 147.7, 437, 135.3, 114.6, 68.4, 52.7, 40.5, 19.1, 15.3; HREIMS [M+H]⁺ *m/z* 167.1055 (calcd for C₁₀H₁₅O₂, 167.1072).

3.5.3.2. Synthesis of 2-oxo-menth-6-en-5 β -ol (F)

Compound F was prepared as previously reported [38]. Briefly, RhCl (PPh₃)₃ (46.2 mg, 0.05 mmol, 5 mol %) was added to a 25-mL RBF containing a stirred solution of E (166 mg, 1.0 mmol, 1.0 equiv) in dry toluene (10 mL) under nitrogen. This flask was sealed with a rubber septum, headspace evacuated, and hydrogen flushed. The reaction mixture was stirred at rt for 14 h. After completion, the solvent was removed under reduced pressure and the residue was purified by flash CC on SiO₂ (5% EtOAc in n-hexane) to afford compound F (153 mg, 91% yield). Oil; $[\alpha]_{\text{D}}^{20} = -62.1$ (c = 0.3, CHCl₃); ¹H NMR (400 MHz, CHCl₃) δ 6.83 (dd, J = 9.0, 1.3 Hz, 1H), 4.31 (d, J = 9.5 Hz, 1H), 2.48 (dd, J = 15.5, 3.6 Hz, 1H), 2.13-2.09 (m, 3H), 2.01-1.87 (m, 1H), 1.75 (s, 3H), 0.95 (d, J = 7.1 Hz, 3H), 0.88 (d, J = 7.0 Hz, 3H); ¹³C NMR (100 MHz, CHCl₃) δ 200.5, 148.8, 135.1, 69.5, 50.8, 37.2, 26.6, 20.5, 16.6, 15.5; HREIMS [M+H]⁺ *m/z* 169.1211 (calcd for C₁₀H₁₇O₂, 169.1229).

3.5.3.3. Synthesis of 2-oxo-menth-6-en-5 β -yl ent-kaurenoate (G)

Compound G was obtained by Steglich esterification [39] from A and F. Briefly, ent-kaurenoic acid (A) (30.2 mg, 0.1 mmol, 1.0 eq), compound F (16.6 mg, 0.1 mmol, 1.0 eq), and dimethylaminopyridine (DMAP) (2.5 mg, 0.02 mmol, 0.2 eq) were mixed within a 10-mL RBF. This flask was sealed with a rubber septum, inner air evacuated, and nitrogen flushed. Anhydrous CH₂Cl₂ (3 mL) was added, followed by 1 M dicyclohexylcarbodiimide (DCC) in CH₂Cl₂ (110 μ L, 0.11 mmol, 1.10 eq). The resulting mixture was stirred overnight then filtered through Celite. The filtrate was concentrated under reduced pressure and the residue was purified by flash CC on SiO₂ (20% EtOAc in n-hexane) to afford compound G (35.7 mg, 79% yield). Oil; $[\alpha]_{\text{D}}^{20} = -58.5$ (c = 0.2, CHCl₃); ¹H NMR (400 MHz, CDCl₃) δ 6.79 (br d, J = 9.4 Hz, 1H), 5.21 (dd, J = 9.4, 1.7 Hz, 1H), 4.78 (br s, 1H),

4.72 (br s, 1H), 2.76-2.68 (m, 2H), 2.65-2.60 (m, 2H), 2.17 (br d, $J = 13.0$ Hz, 1H), 2.08-2.03 (m, 2H), 1.88 (dd, $J = 11.4, 1.1$ Hz, 1H), 1.16 (dd, $J = 11.4, 4.7$ Hz, 1H), 1.91-1.83 (m, 5H), 1.78 (s, 3H), 1.47-1.40 (m, 1H), 1.59-1.51 (m, 6H), 1.16 (s, 3H), 1.08-1.05 (m, 1H), 1.05-1.03 (m, 1H), 1.02 (d, $J = 6.7$ Hz, 3H), 1.01-0.97 (m, 1H), 0.97 (d, $J = 6.7$ Hz, 3H), 0.88 (s, 3H), 0.76 (m, 1H); ^{13}C NMR (100 MHz, CHCl_3) δ 200.6, 177.9, 155.5, 140.1, 134.4, 102.4, 71.5, 57.8, 55.2, 49.5, 45.6, 44.7, 44.4, 43.5, 41.4, 40.2, 39.4, 38.9, 37.7, 37.3, 33.5, 28.6, 26.8, 21.5, 20.3, 19.1, 18.8, 16.7, 15.9, 15.7; HREIMS $[\text{M}+\text{H}]^+$ m/z 453.3345 (calcd for $\text{C}_{30}\text{H}_{45}\text{O}_3$, 453.3369).

3.5.3.4. Synthesis of 2 β -hydroxy-menth-6-en-5 β -yl ent-kaurenoate (**135**)

Compound **135** was obtained from **G**, through Luche reduction using a reported procedure [57]. Briefly, compound **G** (27.1 mg, 0.06 mmol, 1.0 eq), $\text{CeCl}_3 \cdot 7\text{H}_2\text{O}$ (5.6 mg, 0.015 mmol, 0.25 eq), and MeOH (3 mL) was mixed into a 10-mL RBF by stirring at 0 °C. A 1 M NaBH_4 solution (0.06 mL, 0.06 mmol) in MeOH was then added. Reaction mixture was allowed to warm to 20 °C and then stirred at this temperature for 1 h. After completion, reaction was quenched with 2M HCl (2 mL) and extracted with CH_2Cl_2 (3 x 2 mL). The separated CH_2Cl_2 extract was washed with 10% NaCl (2 x 3 mL), dried over MgSO_4 , filtered, and concentrated under reduced pressure. The resulting residue was purified by flash CC on SiO_2 (20% EtOAc in n-hexane) to afford **135** (22.4 mg, 82% yield) (wedelobatin A) [58]. Oil; $[\alpha]_{\text{D}}^{20} -92.6$ (c 0.2, CHCl_3); ^1H NMR (400 MHz, CDCl_3) δ 5.44 (br s, 1H), 5.17 (br d, $J = 8.5$ Hz, 1H), 4.77 (br s, 1H), 4.75 (br s, 1H), 4.03 (t, $J = 3.2$ Hz, 1H), 2.63 (br s, 1H), 2.19 (br d, $J = 13.3$ Hz, 1H), 2.06-2.02 (m, 2H), 1.93 (dd, $J = 11.1, 1.2$ Hz, 1H), 1.14 (dd, $J = 11.1, 5.0$ Hz, 1H), 1.87-1.81 (m, 7H), 1.79 (s, 3H), 1.46-1.41 (m, 2H), 1.60-1.50 (m, 6H), 1.19 (s, 3H), 1.07-1.05 (m, 1H), 1.03-1.01 (m, 1H), 1.00-0.97 (m, 1H), 0.95 (d, $J = 6.8$ Hz, 3H), 0.90 (s, 3H), 0.81 (d, $J = 6.8$ Hz, 3H), 0.78 (m, 1H); ^{13}C NMR (100 MHz, CHCl_3) δ 177.5, 155.4, 139.4, 124.4, 101.9, 71.6, 67.4, 57.2, 55.6, 48.8, 44.2, 43.9, 43.4, 41.1, 41.2, 40.1, 40.1, 39.4, 37.7, 33.0, 30.2, 29.2, 26.2, 22.2, 21.1, 20.2, 19.2, 18.2, 17.2, 16.0; HREIMS $[\text{M}+\text{H}]^+$ m/z 455.3511 (calcd for $\text{C}_{30}\text{H}_{47}\text{O}_3$, 455.3525).

3.5.4. Synthesis of 3 α -cinnamoyloxy-9 β -hydroxy-ent-kaur-16-en-19-oic acid (**301**) 3 α -cinnamoyloxy-ent-kaur-16-en-19-oic acid (**302**), 3 α -p-coumaroyloxy-9 β -hydroxy-ent-kaur-16-en-19-oic acid (**301a**), 3 α -p-coumaroyloxy-ent-kaur-16-en-19-oic acid (**302a**).

Separated reaction, following the same procedure as described for compound **H** (Steglich esterification [39]) of cinnamic acid (**H**) (7.4 mg, 0.05 mmol, 1.0 eq) with **B** (15.9

mg, 0.05 mmol, 1.0 eq) or C (16.7 mg, 0.05 mmol, 1.0 eq) afforded the top-ranked compounds **301** (17.9 mg, 77%) and **302** (17.5 mg, 78%), respectively. Additionally, separated reaction of p-coumaric acid (I) (8.2 mg, 0.05 mmol, 1.0 eq) with B (15.9 mg, 0.5 mmol, 1.0 eq) and C (16.7 mg, 0.5 mmol, 1.0 eq) afforded compounds **301a** (17.0 mg, 71%) and **302a** (16.0 mg, 69%), respectively.

301: Oil; $[\alpha]_{\text{D}}^{20}$ -56.3 (c 0.05, CHCl_3); ^1H NMR (400 MHz, CDCl_3) δ 7.68-7.66 (m, 2H), 7.65 (d, $J = 15.3$ Hz, 1H), 7.40-7.37 (m, 3H), 6.49 (d, $J = 15.3$ Hz, 1H), 4.82 (br s, 1H), 4.77 (br s, 1H), 4.68 (dd, $J = 12.2, 4.5$ Hz, 1H), 2.71 (br d, $J = 14.7$ Hz, 1H), 2.59 (br s, 1H), 2.53-2.50 (m, 1H), 2.28 (dd, $J = 10.5, 1.8$ Hz, 1H), 2.16-2.09 (m, 2H), 2.04-2.00 (m, 1H), 1.93-1.80 (m, 3H), 1.74-1.70 (m, 1H), 1.67-1.55 (m, 3H), 1.51 (dd, $J = 10.5, 5.3$ Hz, 1H), 1.48-1.43 (m, 2H), 1.30-1.27 (m, 1H), 1.26 (s, 3H), 1.12 (s, 3H); ^{13}C NMR (100 MHz, CHCl_3) δ 179.7, 166.7, 157.9, 145.1, 134.7, 130.1, 128.7, 128.2, 118.6, 105.8, 78.5, 75.4, 52.3, 52.1, 49.8, 49.3, 43.7, 42.3, 38.9, 38.5, 34.6, 30.7, 27.3, 25.5, 24.3, 20.5, 17.8; HREIMS $[\text{M}+\text{H}]^+$ m/z 465.2623 (calcd for $\text{C}_{29}\text{H}_{37}\text{O}_5$, 465.2641).

302: Oil; $[\alpha]_{\text{D}}^{20}$ -41.2 (c 0.03, CHCl_3); ^1H NMR (400 MHz, CDCl_3) δ 7.69 (d, $J = 15.1$ Hz, 1H), 7.65-7.62 (m, 2H), 7.47-7.43 (m, 3H), 6.54 (d, $J = 15.1$ Hz, 1H), 4.81 (br s, 1H), 4.75 (br s, 1H), 4.64 (dd, $J = 12.1, 4.7$ Hz, 1H), 2.66 (br s, 1H), 2.36-2.32 (m, 1H), 2.07-2.03 (m, 2H), 1.96 (d, $J = 11.1$ Hz, 1H), 1.93-1.90 (m, 1H), 1.84-1.81 (m, 1H), 1.68-1.62 (m, 3H), 1.55-1.50 (m, 3H), 1.47-1.42 (m, 2H), 1.13-1.07 (m, 2H), 1.05 (br s, 1H), 1.01 (d, $J = 9.3$ Hz, 1H), 1.21 (s, 3H), 0.97 (s, 3H); ^{13}C NMR (100 MHz, CHCl_3) δ 180.2, 166.8, 155.1, 145.3, 134.6, 130.3, 128.8, 128.1, 118.5, 103.1, 79.1, 56.5, 55.1, 48.6, 48.3, 43.8, 43.5, 41.3, 39.6, 39.5, 38.8, 33.3, 24.3, 23.7, 21.5, 18.5, 15.5; HREIMS $[\text{M}+\text{H}]^+$ m/z 449.2678 (calcd for $\text{C}_{29}\text{H}_{37}\text{O}_4$, 449.2692).

301a: Oil; $[\alpha]_{\text{D}}^{20}$ -43.7 (c 0.03, CHCl_3); ^1H NMR (400 MHz, CDCl_3) δ 7.71 (d, $J = 15.0$ Hz, 1H), 7.48 (d, $J = 8.3$ Hz, 2H), 6.91 (d, $J = 8.3$ Hz, 2H), 6.38 (d, $J = 15.0$ Hz, 1H), 4.84 (br s, 1H), 4.71 (br s, 1H), 4.66 (dd, $J = 11.9, 4.9$ Hz, 1H), 2.69 (d, $J = 14.5$ Hz, 1H), 2.57 (br s, 1H), 2.55-2.51 (m, 1H), 2.26 (dd, $J = 10.8, 1.6$ Hz, 1H), 2.20-2.17 (m, 1H), 2.15-2.11 (m, 1H), 2.03-1.98 (m, 1H), 1.91-1.80 (m, 3H), 1.78-1.76 (m, 1H), 1.69-1.63 (m, 2H), 1.58-1.55 (m, 1H), 1.53 (dd, $J = 10.6, 5.1$ Hz, 1H), 1.49-1.44 (m, 2H), 1.29-1.26 (m, 1H), 1.24 (s, 3H), 1.13 (s, 3H); ^{13}C NMR (100 MHz, CHCl_3) δ 181.1, 166.1, 158.1, 156.6, 144.2, 129.1, 127.1, 118.3, 117.9, 107.1, 78.5, 75.4, 52.3, 52.1, 49.6, 49.5, 43.4, 41.8, 38.7, 38.4, 34.5, 30.7, 27.6, 25.7, 24.3, 20.4, 17.9; HREIMS $[\text{M}+\text{H}]^+$ m/z 481.2577 (calcd for $\text{C}_{29}\text{H}_{37}\text{O}_6$, 481.2590).

302a: Oil; $[\alpha]_D^{20}$ -36.5 (c 0.01, CHCl_3); ^1H NMR (400 MHz, CDCl_3) δ 7.76 (d, J = 15.2 Hz, 1H), 7.51 (d, J = 8.1 Hz, 2H), 6.87 (d, J = 8.1 Hz, 2H), 6.45 (d, J = 15.2 Hz, 1H), 4.85 (br s, 1H), 4.77 (br s, 1H), 4.62 (dd, J = 12.0, 4.9 Hz, 1H), 2.67 (br s, 1H), 2.38-2.34 (m, 1H), 2.10-2.06 (m, 2H), 1.97 (d, J = 11.4 Hz, 1H), 1.94-1.90 (m, 1H), 1.85-1.81 (m, 1H), 1.69-1.65 (m, 2H), 1.61-1.53 (m, 4H), 1.49-1.44 (m, 2H), 1.17-1.13 (m, 1H), 1.09-1.07 (m, 1H), 1.04 (br s, 1H), 1.02 (d, J = 9.5 Hz, 1H), 1.21 (s, 3H), 0.97 (s, 3H); ^{13}C NMR (100 MHz, CHCl_3) δ 180.5, 166.3, 158.3, 155.7, 144.8, 129.5, 127.7, 118.5, 117.4, 103.1, 79.1, 56.7, 55.5, 48.7, 48.1, 44.2, 43.8, 41.1, 40.3, 39.3, 39.3, 33.2, 24.5, 23.4, 21.5, 18.7, 15.4; HREIMS $[\text{M}+\text{H}]^+$ m/z 465.2628 (calcd for $\text{C}_{29}\text{H}_{37}\text{O}_5$, 465.2641).

3.6. *LmPTR1* enzyme inhibition assay

Recombinant *LmPTR1* enzyme was obtained, purified, and kinetically characterized, as reported previously [59]. The *in vitro* assessment of selected diterpenes (i.e., **135**, **301**, **302**, **301a**, and **302a**) for *LmPTR1* inhibitory activity was performed through the spectrophotometric monitoring of the enzymatic activity under balanced conditions: *LmPTR1* (30 μg), 7,8-dihydro-L-biopterin (DHB, 20 μM), sodium citrate buffer (20 mM, pH 6.0), 30 $^\circ\text{C}$, and a final assay volume of 600 μL . Each reaction was started by the addition of 250 μM NADPH. Absorbance was monitored at 340 nm (i.e., oxidation of NADPH to NADP^+) for 240 s, and the resulting profile was used to measure the initial reaction rate (IRR) through the respective slope by linear regression.

All recordings were performed in triplicate. PMA was used as the positive control. The resulting IRR values were used to calculate the % inhibition, as $100 - (\text{Ri} / \text{Rc} \times 100)$, where Ri is the IRR in the presence of the inhibitor and Rc is the IRR in the absence of inhibitors (1% DMSO v/v final concentration). The % inhibition for at least five concentrations (range: 0.1–128 μM) for each test compound (diterpenes and PMA) were calculated, and concentration-response curves (% inhibition vs. Log[inhibitor]) were obtained by non-linear regression to determine the IC_{50} using GraphPad Prism 5.0 (GraphPad, San Diego, CA, USA). Finally, K_i^{app} values were calculated using the Cheng-Prusoff equation for competitive inhibition, assuming a 1:1 stoichiometry and that the inhibitor-binding reactions are reversible [41]: $\text{K}_i^{\text{app}} = \text{IC}_{50} / (1 + [\text{S}] / K_m)$, where [S] is the substrate (DHB) concentration and K_m is the Michaelis constant. The substrate K_m was calculated during the kinetic characterization of purified, recombinant *LmPTR1*.

3.7. Hybrid models of *L. braziliensis*, *L. panamensis* and *L. amazonensis*

Hybrid models for *Lb*, *Lp*, and *La*PTR1 were constructed using YASARA software (YASARA (18.4.24) Vienna, Austria: YASARA Biosciences GmbH; 2018) based on the FASTA sequences of *Lb*PTR1 (A4HCP1), *Lp*PTR1 (A0A088SA10), and *La*PTR1 (O09352), which were obtained from the UniProt database (<https://www.uniprot.org/>). The stereochemical qualities of the models were evaluated with PROCHECK [60], which Molecular Diversity evaluated several stereochemical parameters, such as the torsional angles of the main chain, the torsional angles of the side chain, bad contacts or steric impediments, and planarity. PROCHECK generated a Ramachandran graph [45], which verified the allowed and unallowed regions of the main amino acid chain.

The structural quality was evaluated in VERIFY 3D software (<https://services.mbi.ucla.edu/SAVES/>), which analyzes the compatibility of the protein sequence with its 3D structure, according to the chemical environment, and WHAT IF (<https://swift.cmbi.ru.nl/servers/html/index.html>), which analyzes various structural parameters, such as the atomic contacts between residues. The software Discovery Studio Visualizer was used to visualize the modeled protein [61].

3.8. Molecular docking calculations

The *Lm*PTR1 crystal structure (PDB ID: 1E7W), in complex with its respective inhibitor, methotrexate (PDB ID: MTX), was downloaded from PDB [33]. Using Molegro 6.0.1 software, all water compounds were deleted from the enzyme structures, and the enzyme/compound structures were prepared using the same default parameter settings, in the same software package (Score function: MolDock Score; Ligand evaluation: Internal ES, Internal H-Bond, Sp2–Sp2 Torsions, all checked; Number of runs: 10 runs; Algorithm: MolDock SE; Maximum Interactions: 1500; Max. population size: 50; Max. steps: 300; Neighbor distance factor: 1.00; Max. number of poses returned: 5). The docking procedure was performed using a grid with a 15-Å radius and a 0.30-Å resolution to cover the ligand-binding site for the four enzyme structures [14, 20].

The docking procedures for hybrid models of *Leishmania* (*Lb*, *Lp*, and *La*) were performed with the Autodock/Vina (1.1.2) plug-in for PyMOL (1.3r2), under a Python 2.5.2 environment for Windows. Docking calculations were then performed between the minimized ligand through a cube (dimensions 22.5 Å × 22.5 Å × 22.5 Å, grid spacing 0.375 Å) located in the geometric center of the binding pocket (coordinates *Lb*: 18.75, –13.1, 10.25; *Lp*: 18.1, 12.6, 8.0; and *La*: 20.1, 19.6, 7.8), which was identified through

cavities analysis in Molegro 6.0.1. Flexible residues in the binding site were selected for each model. *Lb*: L19, H39, R40, N110, S112, D181, and S227; *Lp*: K17, L19, S112, M179, and I180; and *La*: R18, L19, H38, L188, M233, K244, and Y283. Docking poses were classified according to their docking scores (such as the free energy or affinity). Each calculation was performed in three replicates. Two known PTR1 ligands (DHB and PMA) were used as controls. The two-dimensional (2D)-residual interaction diagrams were visualized on Discovery Studio 2016 Visualizer Client (Biovia, San Diego, CA, USA) [61].

3.9. Molecular dynamics simulations

MD simulations were run in Gromacs 5.0.5 on Ubuntu 12.04 server [62, 63]. Structures **135**, **302**, and **302a** displayed the best poses from docking, and the DHB and PMA structures, as well as the hybrid model of *Lb*PTR1, were employed as the inputs for the MD simulations. The five ligands were prepared by adding hydrogen atoms and the corresponding charges using the AM1-BCC charge scheme in UCSF Chimera. Subsequently, ligand topologies were generated automatically with ACPYPE script. Protein topologies were obtained in Gromacs using the Amber 99SB force field, and the TIP3P water model was implemented. Solvation was performed in a triclinic box using a margin distance of 1.0 nm. The addition of 0.1 M NaCl to complexes and proteins was performed by randomly replacing water molecules until neutrality was achieved [20, 50].

The systems were energy-minimized by 2,000 steps using the steepest descent method. Systems were subjected to NVT equilibration was performed at 310 K for 50 ps, followed by NPT equilibration for 500 ps, using the Parrinello–Rahman method at 1 bar as a reference, using position restraints. Finally, the solute position restraints were released, and a production run for 5 ns was performed. The temperature and pressure were maintained constant at 310 K and 1 bar, respectively. Coordinates were recorded in a 1 fs time step. Electrostatic forces were calculated using the particle-mesh Ewald method. Periodic boundary conditions were used in all simulations, and covalent bond lengths were constrained by the LINCS algorithm. The molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method was used to calculate binding free energies, using the trajectories calculated by the MD simulations [20, 50].

4. Conclusions

Structures **135** and **302** are two kauranes that were identified as hits for anti-leishmanicidal activity, with IC_{50} values against *L. major* below 10 μ M. These two structures were selected from an in-house database comprising 360 kauranes through an *in silico* approach combining machine learning and molecular docking methodologies. Only 5 structures from Asteraceae were classified as active by both methodologies. The *in vitro* results allowed the successful verification of the RF classification model, which predicted that structures **135** and **302** would be active ($pIC_{50} > 5.0$) and that structure **301** would be inactive ($pIC_{50} \geq 5.0$), which was observed experimentally.

Additionally, the inhibitory activity was improved by approximately 60% when a 3 α -*p*-coumaroyloxy group was used in **302** in place of the 3 α -cinnamoyloxy substituent, with **302a** exhibiting a lower Ki^{app} value. Although the tested diterpenes were found to be less active than the positive control, the validity of the designed VS approach for the selection of bioactive molecules against PTR1 was demonstrated, and the computationally-studied binding mode of these selected compounds within the active site of *Lm*PTR1, which causes CL, was explored. These selected compounds can be considered important leads that can be used to obtain more active PTR1 inhibitors.

Finally, because throughout the American continent, other *Leishmania* species are responsible for the clinical diversity of CL and MCL, including *L. amazonensis* (*La*), *L. braziliensis* (*Lb*), and *L. panamensis* (*Lp*), molecular docking calculations and MD simulations were performed for the entire set of kauranes (including **301a** and **302a**), and the compounds **135**, **302**, and **302a** were identified as potential multispecies agents. Therefore, this study describes a valuable screening approach for the identification of lead compounds in natural products, which can contribute to the further development of alternative chemotherapies against this group of diseases.

References

- [1]. Utzinger, J.; Becker, S.L.; Knopp, S.; Blum, J.; Neumayr, A.L.; Keiser, J.; Hatz, C.F. Neglected tropical diseases: diagnosis, clinical management, treatment and control. *Swiss medical weekly: official journal of the Swiss Society of Infectious Diseases, the Swiss Society of Internal Medicine, the Swiss Society of Pneumology* **2012**, 142.
- [2]. López-Arencibia, A.; Bethencourt-Estrella, C.J.; Freijo, M.B.; Reyes-Batlle, M.; Sifaoui, I.; San Nicolás-Hernández, D.; McNaughton-Smith, G.; Lorenzo-Morales, J.;

- Abad-Grillo, T.; Piñero, J.E. New phenalenone analogues with improved activity against *Leishmania* species. *Biomedicine & Pharmacotherapy* **2020**, *132*, 110814.
- [3].Organization, W.H. Leishmaniasis. Available online: (accessed on 11 July 2020).
- [4].Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M.D.F.; Scotti, M.T. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. *Molecules* **2017**, *22*, 79.
- [5].De Brito, R.C.F.; de Oliveira Aguiar-Soares, R.D.; de Oliveira Cardoso, J.M.; Coura-Vital, W.; Roatt, B.M.; Reis, A.B. Recent advances and new strategies in Leishmaniasis diagnosis. *Applied Microbiology and Biotechnology* **2020**, 1-12.
- [6].Akilov, O.E.; Khachemoune, A.; Hasan, T. Clinical manifestations and classification of Old World cutaneous leishmaniasis. *International journal of dermatology* **2007**, *46*, 132-142.
- [7].Ghatee, M.A.; Taylor, W.R.; Karamian, M. The Geographical Distribution of Cutaneous Leishmaniasis Causative Agents in Iran and Its Neighboring Countries, A Review. *Frontiers in Public Health* **2020**, *8*.
- [8].Paz, C.; Samake S Fau - Anderson, J.M.; Anderson Jm Fau - Faye, O.; Faye O Fau - Traore, P.; Traore P Fau - Tall, K.; Tall K Fau - Cisse, M.; Cisse M Fau - Keita, S.; Keita S Fau - Valenzuela, J.G.; Valenzuela Jg Fau - Doumbia, S.; Doumbia, S. *Leishmania major*, the predominant *Leishmania* species responsible for cutaneous leishmaniasis in Mali.
- [9].Vojtkova, B.; Spitzova, T.; Votypka, J.; Lestinova, T.; Kominkova, I.; Hajkova, M.; Santos-Mateus, D.; Miles, M.A.; Volf, P.; Sadlova, J. Central Asian Rodents as Model Animals for *Leishmania major* and *Leishmania donovani* Research. *Microorganisms* **2020**, *8*, 1440.
- [10].Sánchez-Suárez, J.; Bernal, F.A.; Coy-Barrera, E. Colombian Contributions Fighting Leishmaniasis: A Systematic Review on Antileishmanials Combined with Chemoinformatics Analysis. *Molecules* **2020**, *25*, 5704.
- [11].Anversa, L.; Tiburcio, M.G.S.; Richini-Pereira, V.B.; Ramirez, L.E. Human leishmaniasis in Brazil: a general review. *Revista da Associação Médica Brasileira* **2018**, *64*, 281-289.
- [12].Gervazoni, L.F.O.; Barcellos, G.B.; Ferreira-Paes, T.; Almeida-Amaral, E.E. Use of natural products in leishmaniasis chemotherapy: an overview. *Frontiers in Chemistry* **2020**, *8*, 1031.

- [13].Rodrigues, B.C.; Ferreira, M.F.; Barroso, D.H.; da Motta, J.O.C.; de Paula, C.D.R.; Porto, C.; Martins, S.S.; Gomes, C.M.; Sampaio, R.N.R. A retrospective cohort study of the effectiveness and adverse events of intralesional pentavalent antimonials in the treatment of cutaneous leishmaniasis. *International Journal for Parasitology: Drugs and Drug Resistance* **2020**, *14*, 257-263.
- [14].Acevedo, C.H.; Scotti, L.; Scotti, M.T. In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018**, *13*, 634-645.
- [15].Varela, M.T.; Fernandes, J.P.S. Natural products: Key prototypes to drug discovery against neglected diseases caused by trypanosomatids. *Current Medicinal Chemistry* **2020**, *27*, 2133-2146.
- [16].Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products* **2020**, *83*, 770-803.
- [17].Herrera-Acevedo, C.; Perdomo-Madrigal, C.; Muratov, E.N.; Scotti, L.; Scotti, M.T. Discovery of Alternative Chemotherapy Options for Leishmaniasis via Computational Studies of Asteraceae. *ChemMedChem* **2020**.
- [18].Scotti, M.T.; Herrera-Acevedo, C.; Oliveira, T.B.; Costa, R.P.O.; Santos, S.Y.K.d.O.; Rodrigues, R.P.; Scotti, L.; Da-Costa, F.B. SistemX, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules* **2018**, *23*, 103.
- [19].Bernal, F.A.; Coy-Barrera, E. In-silico analyses of sesquiterpene-related compounds on selected Leishmania enzyme-based targets. *Molecules* **2014**, *19*, 5550-5569.
- [20].Herrera-Acevedo, C.; Maia, M.D.S.; Cavalcanti, É.B.V.S.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Selection of antileishmanial sesquiterpene lactones from SistemX database using a combined ligand-/structure-based virtual screening approach. *Molecular Diversity* **2020**, 1-17.
- [21].Melo, T.S.; Gattass, C.R.; Soares, D.C.; Cunha, M.R.; Ferreira, C.; Tavares, M.T.; Saraiva, E.; Parise-Filho, R.; Braden, H.; Delorenzi, J.C. Oleanolic acid (OA) as an antileishmanial agent: Biological evaluation and in silico mechanistic insights. *Parasitology international* **2016**, *65*, 227-237.
- [22].Shah, S.M.; Ullah, F.; Ayaz, M.; Sadiq, A.; Hussain, S.; Shah, S.A.A.; Nadhman, A. β -Sitosterol from *Ifloga spicata* (Forssk.) Sch. Bip. as potential anti-leishmanial agent against leishmania tropica: docking and molecular insights. *Steroids* **2019**, *148*, 56-62.

- [23].Seaman, F.; Bohlmann, F.; Zdero, C.; Mabry, T.J. *Diterpenes of flowering plants: Compositae (Asteraceae)*; Springer Science & Business Media: 2012.
- [24].Nichol, C.A.; Smith, G.K.; Duch, D.S. BIOSYNTHESIS AND METABOLISM OF TETRAHYDROBIOPTERIN AND MOLYBDOPTERIN. *Annual Review of Biochemistry* **1985**, *54*, 729-764, doi:10.1146/annurev.bi.54.070185.003501.
- [25].Cruciani, G.; Crivori, P.; Carrupt, P.A.; Testa, B. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *Journal of Molecular Structure: THEOCHEM* **2000**, *503*, 17-30.
- [26].Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences* **2000**, *11*, S29-S39.
- [27].Berthold, M.R.; Cebon, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter* **2009**, *11*, 26-31.
- [28].Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S.Y.; Mumper, R.J.; Tropsha, A. Quantitative nanostructure– activity relationship modeling. *ACS nano* **2010**, *4*, 5703-5712.
- [29].Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29-36.
- [30].Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* **1975**, *405*, 442-451.
- [31].Cai, C.; Zhang, Y.; Yang, D.; Hao, X.; Li, S. Two new kaurane-type diterpenoids from *Wedelia chinensis* (Osbeck.) Merr. *Natural product research* **2017**, *31*, 2531-2536.
- [32].Li, S.-F.; Ding, J.-Y.; Li, Y.-T.; Hao, X.-J.; Li, S.-L. Antimicrobial diterpenoids of *Wedelia trilobata* (L.) Hitchc. *Molecules* **2016**, *21*, 457.
- [33].Gourley, D.G.; Schüttelkopf, A.W.; Leonard, G.A.; Luba, J.; Hardy, L.W.; Beverley, S.M.; Hunter, W.N. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. *Nature structural biology* **2001**, *8*, 521-525.
- [34].Gobu, F.-R.; Chen, J.-J.; Zeng, J.; Wei, W.-J.; Wang, W.-F.; Lin, C.-J.; Gao, K. Isolation, structure elucidation, and immunosuppressive activity of diterpenoids from *Ligularia fischeri*. *Journal of Natural Products* **2017**, *80*, 2263-2268.

- [35].Moreira, I.C.; Roque, N.F.; Vilegas, W.; Zalewski, C.A.; Lago, J.H.G.; Funasaki, M. Genus *Xylopia* (Annonaceae): chemical and biological aspects. *Chemistry & biodiversity* **2013**, *10*, 1921-1943.
- [36].Silva, D.M.; Costa, E.V.; Nogueira, P.C.d.L.; Moraes, V.R.d.S.; Cavalcanti, S.C.d.H.; Salvador, M.J.; Ribeiro, L.H.G.; Gadelha, F.R.; Barison, A.; Ferreira, A.G. Ent-kaurane diterpenoids and other constituents from the stem of *Xylopia laevigata* (Annonaceae). *Química Nova* **2012**, *35*, 1570-1576.
- [37].Wu, Y.-C.; Hung, Y.-C.; Chang, F.-R.; Cosentino, M.; Wang, H.-K.; Lee, K.-H. Identification of ent-16 β , 17-dihydroxykauran-19-oic acid as an anti-HIV principle and isolation of the new diterpenoids annosquamosins A and B from *Annona squamosa*. *Journal of natural products* **1996**, *59*, 635-637.
- [38].Zhang, P.; Li, Y.; Yan, Z.; Gong, J.; Yang, Z. Asymmetric Total Synthesis of (-)-Pavidolide B via a Thiyl-Radical-Mediated [3+ 2] Annulation Reaction. *The Journal of Organic Chemistry* **2019**, *84*, 15958-15971.
- [39].Neises, B.; Steglich, W. Simple method for the esterification of carboxylic acids. *Angewandte Chemie International Edition in English* **1978**, *17*, 522-524.
- [40].Šťastná, E.; Černý, I.; Pouzar, V.; Chodounská, H. Stereoselectivity of sodium borohydride reduction of saturated steroidal ketones utilizing conditions of Luche reduction. *Steroids* **2010**, *75*, 721-725.
- [41].Cheng, Y.C.; Prusoff, W.H. the concentration of inhibitor which causes 50 percent inhibition (I) of an enzymatic reaction. *Biochem. Pharmacol* **1973**, *22*, 3099-3108.
- [42].Jeronimo, S.M.B.; De Queiroz Sousa, A.; Pearson, R.D. Chapter 94 - Leishmaniasis. In *Tropical Infectious Diseases (Second Edition)*, Guerrant, R.L., Walker, D.H., Weller, P.F., Eds.; Churchill Livingstone: Philadelphia, 2006; pp. 1095-1113.
- [43].Lima, B.S.d.S.; Esteves, B.B.; Fialho-Júnior, L.C.; Mendes, T.A.d.O.; Pires, S.d.F.; Chapeourouge, A.; Perales, J.; de Andrade, H.M. Study of the differentially abundant proteins among *Leishmania amazonensis*, *L. braziliensis*, and *L. infantum*. *Plos one* **2020**, *15*, e0240612.
- [44].Krieger, E.; Vriend, G. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981-2982.
- [45].Lovell, S.C.; Davis, I.W.; Arendall Iii, W.B.; De Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics* **2003**, *50*, 437-450.

- [46].Krieger, E.; Hooft, R.W.W.; Nabuurs, S.; Vriend, G. PDBFinderII—a database for protein structure analysis and prediction. **2004**.
- [47].Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function, and Bioinformatics* **2009**, 77, 114-122.
- [48].Salgado-Almario, J.; Hernández, C.A.; Ovalle-Bracho, C. Geographical distribution of Leishmania species in Colombia, 1985-2017. *Biomédica* **2019**, 39, 278-290.
- [49].Belli, A.A.; Miles, M.A.; Kelly, J.M. A putative Leishmania panamensis/Leishmania braziliensis hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. *PARASITOLOGY-CAMBRIDGE*- **1994**, 109, 435-435.
- [50].Rozo-Lugo, C.; Cuca-Suárez, L.E.; Schmidt, T.J.; Coy-Barrera, E. Tetrahydrobenzofuran-6 (2 H)-one Neolignans from Ocotea heterochroma: Their Platelet Activating Factor (PAF) Antagonistic Activity and in Silico Insights into the PAF Receptor Binding Mode. *Journal of natural products* **2018**, 81, 1968-1975.
- [51].Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **1990**, 215, 403-410.
- [52].Noble, W.S. What is a support vector machine?
- [53].Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R. QSAR modeling: where have you been? Where are you going to? *Journal of medicinal chemistry* **2014**, 57, 4977-5010.
- [54].Fourches, D.; Muratov, E.; Tropsha, A. Curation of chemogenomics data. *Nature chemical biology* **2015**, 11, 535-535.
- [55].Muratov, E.N.; Varlamova, E.V.; Artemenko, A.G.; Polishchuk, P.G.; Kuz'min, V.E. Existing and developing approaches for QSAR analysis of mixtures. *Molecular informatics* **2012**, 31, 202-221.
- [56].Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of medicinal chemistry* **2010**, 53, 2719-2740.
- [57].Fruhmann, P.; Hametner, C.; Mikula, H.; Adam, G.; Krska, R.; Fröhlich, J. Stereoselective luche reduction of deoxynivalenol and three of its acetylated derivatives at C8. *Toxins* **2014**, 6, 325-336.

- [58].Ma, B.-J.; Wen, C.-N.; Gao, Y.; Ren, F.-C.; Wang, F.; Liu, J.-K. ent-Kaurane diterpenoids from the plant *Wedelia trilobata*. *Natural products and bioprospecting* **2013**, 3, 107-111.
- [59].Borsari, C.; Luciani, R.; Pozzi, C.; Poehner, I.; Henrich, S.; Trande, M.; Cordeiro-da-Silva, A.; Santarem, N.; Baptista, C.; Tait, A. Profiling of flavonol derivatives for the development of antitrypanosomatidic drugs. *Journal of medicinal chemistry* **2016**, 59, 7598-7616.
- [60].Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **1993**, 26, 283-291.
- [61].Biovia, D.S. Discovery Studio Modeling Environment, Release 2017, San Diego: DassaultSystèmes, 2016. Available from:(Accessed 1 September 2016) **2016**.
- [62].Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1, 19-25.
- [63].Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer physics communications* **1995**, 91, 43-56.

The background of the page is a complex, abstract pattern. It features a network of interconnected nodes and lines, resembling a molecular structure or a data network. The nodes are represented by small circles in various shades of blue and grey, while the lines are thin and light grey. Interspersed within this network are several hexagonal shapes, some of which are solid blue and others are outlined in blue. The overall effect is a dense, layered composition that suggests themes of science, technology, or interconnectedness.

Parte B

Kaurane-Type Diterpenoids as Potential Inhibitors of Dihydrofolate Reductase-Thymidylate Synthase in New World *Leishmania* Species

Chonny Herrera-Acevedo^{1,2}, Renata Priscila Barros de Menezes¹ Natália Ferreira de Sousa¹, Luciana Scotti¹, Marcus Tullius Scotti¹ and Ericsson Coy-Barrera³, *

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil; chonny622@gmail.com; renatabarros@ltf.ufpb.br; nferreiradesousa.nfs@gmail.com; luciana.scotti@gmail.com and mtscotti@gmail.com.

² Department of Chemical Engineering, Universidad ECCI, Carrera 19# 49-20, 111311, Bogotá D.C., Colombia

³ Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia; ericsson.coy@unimilitar.edu.co

*Correspondence: ericsson.coy@unimilitar.edu.co

Artigo publicado em Antibiotics (IF: 4.80)

DOI Number: <https://doi.org/10.3390/antibiotics12040663>

Ano de publicação: 2023

Abstract: The bifunctional enzyme Dihydrofolate reductase-thymidylate synthase (DHFR-TS) plays a crucial role in the survival of the *Leishmania* parasite, as folates are essential cofactors for purine and pyrimidine nucleotide biosynthesis. However, DHFR inhibitors are largely ineffective in controlling trypanosomatid infections, largely due to the presence of Pteridine reductase 1 (PTR1). Therefore, the search for structures with dual inhibitory activity against PTR1/DHFR-TS is crucial in the development of new anti-*Leishmania* chemotherapies. In this research, using the *Leishmania major* DHFR-TS recombinant protein, enzymatic inhibitory assays were performed on four kauranes and two derivatives that were previously tested against *Lm*PTR1. The structure **302** (6.3 μ M) and its derivative **302a** (4.5 μ M) showed the lowest IC₅₀ values among the evaluated molecules. To evaluate the mechanism of action of these structures, molecular docking calculations and molecular

dynamics simulations were performed using a DHFR-TS hybrid model. Results showed that hydrogen bond interactions are critical for the inhibitory activity against *Lm*DHFR-TS, as well as the presence of the *p*-hydroxyl group of the phenylpropanoid moiety of **302a**. Finally, additional computational studies were performed on DHFR-TS structures from *Leishmania* species that cause cutaneous and mucocutaneous leishmaniasis in the New World (*L. braziliensis*, *L. panamensis*, and *L. amazonensis*) to explore the targeting potential of these kauranes in these species. It was demonstrated that structures **302** and **302a** are multi-*Leishmania* species compounds with dual DHFR-TS/PTR1 inhibitory activity.

Keywords: kauranes; *Leishmania*; Asteraceae; machine learning; DHFR-TS; diterpenes; natural products.

1. Introduction

Leishmaniasis is a neglected tropical disease (NTD) caused by *Leishmania* parasites, a type of trypanosomatid protozoa [1]. The disease affects 15 million people globally, presenting in three forms: cutaneous (CL), mucocutaneous (ML), and visceral (VL) [2,3]. Despite the public health concerns and need for control, current treatments, including pentavalent antimony salts as the first-line drugs or amphotericin B, pentamidine, miltefosine, or paromomycin as second-line drugs, are frequently toxic, expensive, and only marginally effective with increasing resistance outbreaks [3-5]. Although attempts to discover more effective and safe alternatives through drug discovery [1,2,6], limited progress has been made, making the search for new antileishmanial chemotherapies necessary [7].

A metabolic pathway that is traditionally considered a crucial target against trypanosomatid parasites involves the inhibition of dihydrofolate reductase (DHFR) in the biosynthesis of folate-like cofactors [8]. DHFR (EC 1.5.1.3) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolates (H₂Fs) to 5,6,7,8-tetrahydrofolates (H₄Fs) [9], which are necessary for maintaining adequate intracellular folate concentrations [8,9]. In trypanosomatids, a single, fused gene encodes a bifunctional enzyme that has both the DHFR domain and the thymidylate synthase (TS) domain [10]. This bifunctional enzyme is crucial for the parasite's survival because folates are essential cofactors for the biosynthesis of purine and pyrimidine nucleotides. As a result, inhibition of this single polypeptide can affect two steps of this essential pathway [11]. In contrast, humans have

separate mono-functional polypeptides for DHFR and TS, leading to structural differences and unique roles in human folate production [8]. This makes the DHFR-TS combination an attractive molecular target for the development of antimicrobial agents. In fact, antifolate-based antimicrobial drugs such as methotrexate (MTX), trimethoprim, and pyrimethamine are already in use [8,11].

However, *Leishmania* parasites are auxotrophic for folate, meaning they have a sophisticated metabolic pathway for acquiring folate from the host and incorporating it into intermediate or alternative metabolism through the action of pteridine reductase (PTR1) [12]. PTR1 (EC 1.5.1.33) transforms conjugated and non-conjugated pterins, including the reduction of biopterin to dihydrobiopterin, and then to tetrahydrobiopterin. This catalytic role is crucial for maintaining vital intracellular levels of tetrahydropterin and has been shown to be an essential component of growth *in vivo* through gene expression studies [13]. Since PTR1 is less sensitive to the effect of MTX and catalyzes folate reduction, this explains the therapeutic failures of antifolate drugs against trypanosomatid parasites [12,14,15]. Thus, an appropriate strategy would involve searching for dual inhibitors of PTR1 and DHFR as antileishmanial agents [16], and natural compounds are still considered a vast source of bioactive agents [2].

In this context, a class of bioactive naturally-occurring compounds known as Kaurane-type diterpenes has been shown to exhibit antileishmanial activity at various levels [17-19]. Based on this evidence, a previous *in silico* and *in vitro* study was performed on a custom-made library of 360 compounds to select Kaurane-type diterpenes against *Leishmania major* PTR1 (*LmPTR1*). The top-ranked compounds and two semi-synthetic derivatives were found to have half-maximal inhibitory concentrations (IC₅₀) less than 10 µg/mL. Given these results and with the aim of exploring dual inhibitors of DHFR/PTR1, the present study investigated the selection of kauranes with activity against *L. major* DHFR-TS.

2. Results and discussion

2.1. Kauranes 302 and its derivative 302a have dual *in vitro* enzymatic activity against *L. major* PTR1/DHFR-TS

The potential dual enzymatic activity of *L. major* PTR1/DHFR-TS for the diterpene esters **135**, **301**, **302**, **301a**, and **302a** (which have already been evaluated against *L. major* PTR1 [20], as shown in Figure 1a), along with structure **4**, which was synthesized from the

kaurane **148** (ent-kaurane-3-oxo-16 α ,17-diol, shown in Figure 1b), a compound isolated from *Euphorbia gracilis* Jacq. (Euphorbiaceae), was evaluated using spectrophotometric monitoring of enzymatic activity under a standard DHFR assay. This was done with a range of test compound concentrations (0.1–128 μ M), and methotrexate was used as a positive control.

The IC₅₀ values were calculated based on the concentration-response behavior within the range of 0.1–128 μ M, resulting in values ranging from 4.5 to 11.2 μ M (pIC₅₀ values ranging from 4.95 to 5.35). Then, using the Cheng–Prusoff equation and assuming reversible competitive inhibition and a 1:1 stoichiometry [21], the apparent inhibitory constant (K_i^{app}) was calculated for the selected kauranes using the IC₅₀ results, as shown in Table 1.

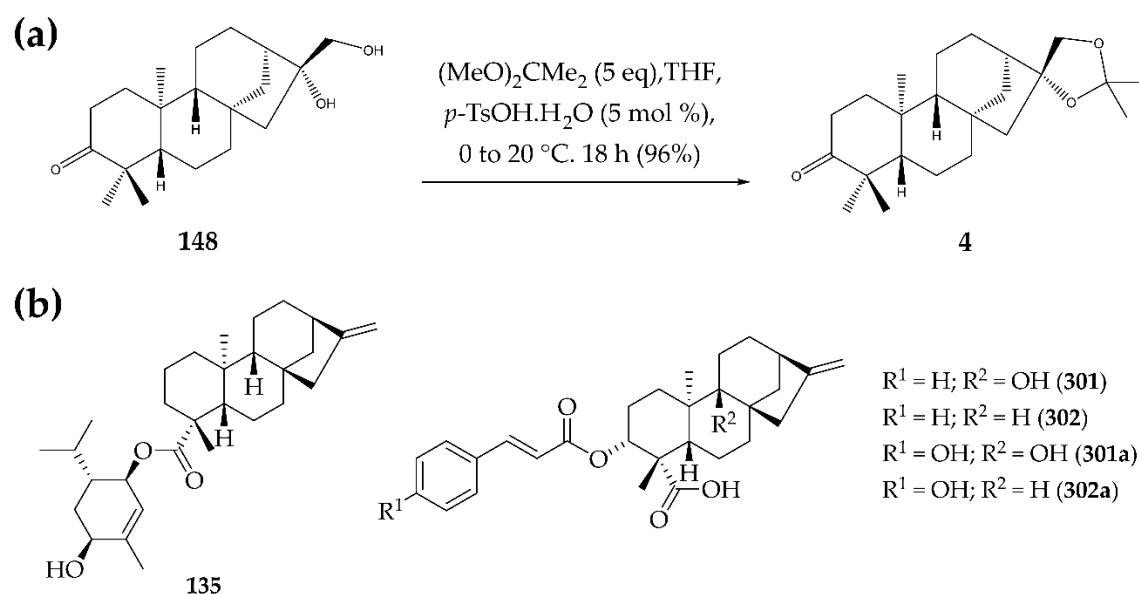


Figure 1. (a) Synthesis of compound **4**. (b) Structures of selected kaurane-type diterpenes (**135**, **301**, **302**) and their derivatives (**301a** and **302a**)

Table 1. Results of enzymatic activity against *L. major* dihydrofolate reductase (*LmDHFR*) for selected kaurane-type diterpenes.

Compound	4	135	302	301	302a	301a	MTX
IC ₅₀ (μ M)	7.6	11.2	6.3	8.8	4.5	7.9	1.4
Confidence Interval (95%)	6.9–8.1	10.2–12.1	5.8–6.9	8.0–9.9	3.9–5.2	7.1–8.4	1.1–1.8
K_i^{app}	0.81	1.20	0.68	0.94	0.48	0.85	0.15

The evaluated structures showed similar IC₅₀ values. Among the six tested diterpenes, structure **135** was the least active, which was contrary to what was observed with PTR1. Structure **301**, which was classified as inactive against PTR1, showed a different behavior against DHFR-TS with a pIC₅₀ value above 5.0, and was classified as active against this enzyme, according to the cutoff value used to build the machine learning model of *L. major* (pIC₅₀ = -log IC₅₀) [20].

Using MolpredictX, a recent web tool developed in the Laboratory of Cheminformatics at the Federal University of Paraíba, which provides predictions for 27 different biological activities, including *L. major*, the structures **301** and **301a** were classified as active. This tool provides qualitative predictions of molecule activity (active or inactive) and a quantitative probability of activity based on molecular descriptors. [22].

For DHFR-TS, the kaurane-type diterpenes **301**, **302**, **301a**, and **302a** showed similar pIC₅₀ values above 5.0, indicating that the 9-hydroxyl group at the diterpene moiety is not relevant to the inhibitory activity as observed with PTR1, and suggesting different mechanisms of action for these two enzymes in *Leishmania*. Additionally, the *p*-hydroxyl group has a favorable influence on the inhibitory activity of the evaluated kauranes, reducing the inhibitory constant (K_{iapp}) values by 10-30% for **301a** and **302a**, respectively, compared to the kauranes **301** and **302**, which do not have this hydroxyl group present in their structures.

The structures **302** and **302a** showed the lowest K_{iapp} values among the six tested structures against DHFR-TS (despite both having K_{iapp} values that are higher than MTX). These two structures also displayed a similar behavior with lower K_{iapp} values in previous enzymatic assays against *L. major* PTR1 [20], which indicates that these two structures have dual *in vitro* enzymatic activity against *L. major* PTR1/DHFR-TS, with the 9-hydroxyl group at the diterpene moiety being the critical structural feature for the observed dual action against these targets.

2.2. Hybrid Model of *L. major* DHFR-TS and Molecular Docking Calculations.

To examine the mechanism of action of the tested kauranes and determine whether the kauranes that previously showed inhibitory activity against pteridine reductase 1 (PTR1) also act against dihydrofolate reductase-thymidylate synthase (DHFR-TS), a molecular docking study was conducted using a *Lm*DHFR-TS hybrid model built in the YASARA software (YASARA Biosciences GmbH, Vienna, Austria; 2018). The model's reliability and stereochemical qualities were evaluated through Ramachandran, WHAT IF,

and VERIFY 3D plots, as well as Z-scores of dihedrals, which describe the deviation of the model's quality from the average high-resolution X-ray structure. The Ramachandran plot showed that 96.9 of residues were in the most favored regions, with 99.5% in allowed regions and only 0.5% (corresponding to five amino acids) in the outlier region, indicating that the *LmDHFR-TS* model was satisfactory (Supplementary Material).

The VERIFY 3D (<https://services.mbi.ucla.edu/SAVES/>) results showed that 92.6% of residues had an averaged 3D-1D score of ≥ 0.2 , indicating a reliable model. The Coarse Packing Quality Control of the *LmDHFR-TS* model, evaluated using WHAT IF, showed a mean score of -0.594, with only 1.7% of residues (8 of 520 amino acids) scoring -5.0 or lower. The dihedral quality was classified as optimal for the *LmDHFR-TS* hybrid model, with values above 1.085 [24].

Molecular docking calculations for the selected kaurane dataset and derivatives **301a** and **302a** were performed using Molegro 6.0 software and the previously validated *L. major* DHFR-TS hybrid model. The docking energy values ranged from -62.85 to -81.43 kJ/mol, with all structures showing higher values than the positive control MTX (-107.60 kJ/mol). Interestingly, kaurane **302** (-76.53 kJ/mol) and its derivative **302a** (-81.43 kJ/mol), which showed the highest inhibitory activity against *L. major* DHFR-TS in the enzymatic assay, had the lowest docking scores among the evaluated molecules. The analysis of the docking conformations revealed that the phenylpropanoid moiety of these two molecules adopted a similar conformation in the active site of *L. major* DHFR-TS, with the p-hydroxyl group identified as a crucial feature for the observed inhibitory activity (Figure 2h).

Table 2. Docking energies of six tested structures and MTX for *L. major* DHFR-TS. SD = standard deviation; RMSD values = root mean square deviation.

Structure	Docking score (kJ/mol)	RMSD (Å)	SD
4	-70.25	0.68	5.7
135	-62.85	1.23	8.6
301	-73.34	1.09	10.3
302	-76.53	1.13	4.9
301a	-72.26	0.89	9.8
302a	-81.43	1.29	6.4
MTX	-107.60	0.24	5.9

Using a two-dimensional analysis, critical interactions with active site amino acid residues of the enzyme were identified. It was observed that hydrogen bond interactions are directly related to the IC₅₀ values obtained in the enzymatic assay. Structures **302** and **302a**, which had the lowest IC₅₀ values, showed two hydrogen bond interactions involving residues I45 and S86 for **302** and W47 for **302a**, and the carbon-19 of these two kauranes (Figure 2d-f). The interaction with residue S86 was also observed for structure **301**, which had a moderate IC₅₀ value among the tested structures, with the only hydrogen bond interaction being observed in this kaurane (Figure 2c). This behavior was also observed in structure **4** (Figure 2a), which only interacted with residue A32 through hydrogen bonds.

The positive control, MTX, showed three hydrogen bond interactions with residues D52, K57, and V30. The interaction with V30 potentially having a crucial role in the inhibition of DHFR-TS. This interaction was only observed in the derivative **302a**, which had the highest inhibitory activity among the tested molecules. Interestingly, this hydrogen bond interaction was formed with the *p*-hydroxyl group of the phenylpropanoid moiety of **302a**, reinforcing the importance of this structural feature in the dual *L. major* PTR1/DHFR-TS inhibitory activity.

Residue F56 also plays a key role in the inhibition of *L. major* DHFR-TS when it interacts with aromatic regions in the kaurane series, as the most active molecules exhibited a π - π interaction between the phenyl group of the amino acid and the pteridine ring and phenylpropanoid moiety of MTX and structure **302a**, respectively. A different behavior was observed for structure **302** and the derivative **301a**, which had intermediate inhibitory activity against DHFR-TS. These two molecules, along with structure **135**, showed a π - σ interaction with the hydrogens of the kaurane region. Structure **302a** was the only structure that showed an unfavorable interaction with residue M53, which is important for MTX, with a π -sulfur interaction being established (Figure 2f-g).

2.3. Kaurane 302 and its derivative 302a may have the potential to inhibit DHFR-TS in different species of *Leishmania* from the New World.

Leishmaniasis contracted in North and South America is referred to as "new world leishmaniasis" [25]. Studying this type of species is crucial for the control and elimination of the disease, as there is a high diversity of *Leishmania* species in the Americas, with high concentrations of different species found in countries such as Brazil and Colombia, leading to a significant disease burden [26]. Some of the main new world *Leishmania* species include: *Leishmania panamensis*, which is the primary cause of cutaneous leishmaniasis

(CL) in Panama and has been found to infect both anthropophilic vectors and mammalian reservoirs [27]; *Leishmania braziliensis*, a pathogenic agent of CL and mucocutaneous leishmaniasis (MCL), primarily distributed in South and Central America [25-28]; and *Leishmania amazonensis*, an etiological agent of diffuse CL and tegumentary leishmaniasis (TL) [29]. In previous research, molecular docking calculations and MD simulations using PTR1 hybrid models of *L. braziliensis*, *L. amazonensis*, and *L. panamensis* have identified the kauranes **135**, **302**, and its derivative **302a** as potential multi-species agents [20].

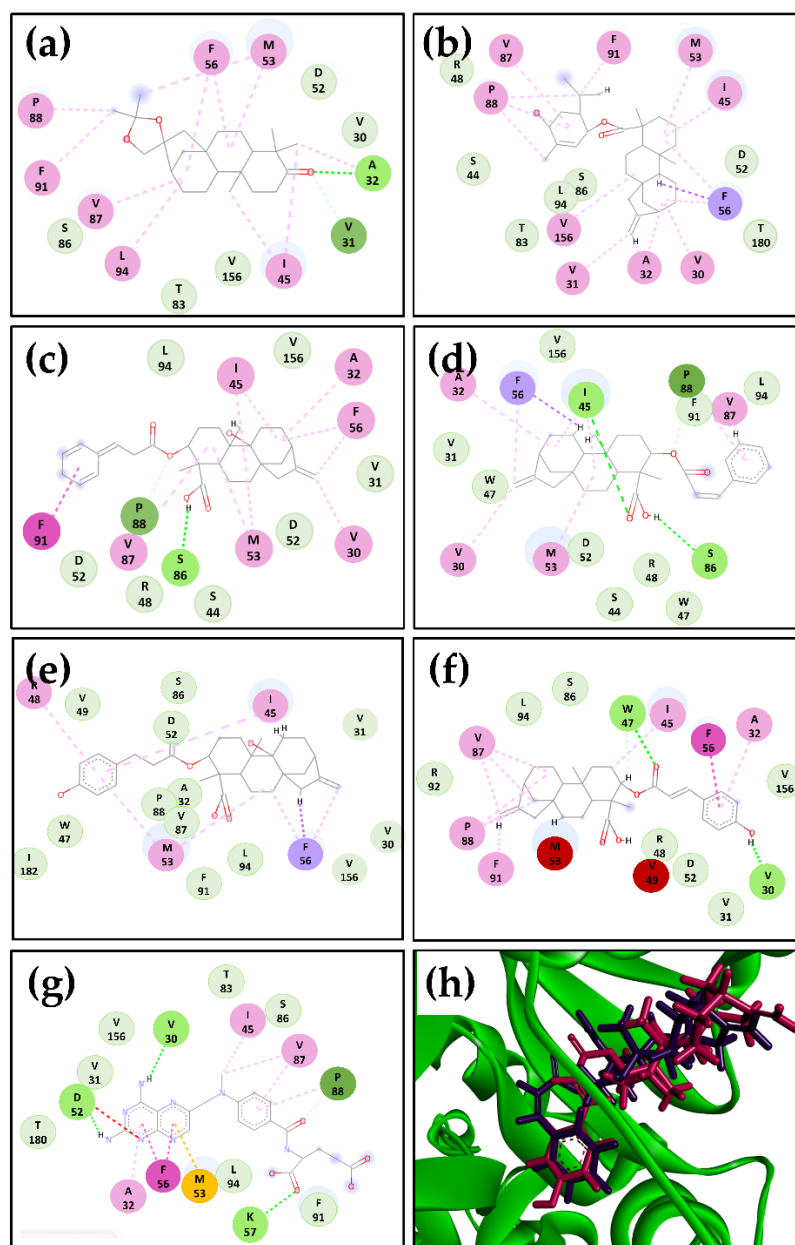


Figure 2. Two-dimensional residual interaction diagrams of (a) Structure **4**, (b) structure **135**, (c) Structure **301**, (d) Structure **302**, (e) Structure **301a**, (f) Structure **302a** and (g) Methotrexate (MTX). Interacting residues are shown as colored circles depending on the

interactions (as colored dashed lines): H-bond (lime), Van der Waals (green), π - σ (purple), π -alkyl (pink), π - π (fuchsia), unfavorable (red), and carbon H-bond (teal) interactions. (h) Docking conformations of structure **302** (purple) and its derivative **302a** (pink) in the active site of *L. major* DHFR-TS (green).

To evaluate their potential dual inhibitory activity against PTR1 and DHFR-TS, hybrid models of DHFR-TS for these three *Leishmania* species were built, and molecular docking calculations and MD simulations were performed using the four kauranes and two derivatives, which were previously tested against the DHFR-TS recombinant. The Ramachandran plot of these three hybrid models showed that the main possible chain conformations included more than 97.2% of residues in the most favored regions for the three hybrid models, with 99.7% of residues in allowed regions. All models showed three residues (0.3%) in disallowed regions (outliers; Supplementary Material).

Table 3. The VINA score values for six tested structures and MTX (methotrexate) for dihydrofolate reductase-thymidylate synthase (DHFR-TS) of *Leishmania braziliensis*, *Leishmania panamensis*, and *Leishmania amazonensis*. SD = standard deviation; RMSD values = root mean square deviation.

Structure	<i>L. braziliensis</i>			<i>L. panamensis</i>			<i>L. amazonensis</i>		
	VINA Score (kcal/mol)	SD	RMSD	VINA Score (kcal/mol)	SD	RMSD	VINA Score (kcal/mol)	SD	RMSD
4	-10.70	0.05	0.13	-10.96	0.07	0.46	-10.68	0.04	0.11
135	-10.50	0	0.21	-10.19	0.03	0.31	-10.52	0.06	0.25
302	-10.90	0.05	0.55	-10.44	0.05	2.73	-10.55	0.15	0.86
302^a	-11.17	0.13	0.61	-12.55	0.28	0.56	-10.60	0.08	0.61
301	-10.40	0.10	0.92	-10.84	0.07	1.68	-10.85	0.05	0.64
301^a	-10.66	0.20	0.45	-12.54	0.08	0.86	-11.14	0.15	0.79
MTX	-9.64	0.07	1.87	-9.45	0.15	1.48	-9.54	0.07	1.71

The analysis of the docking results showed that for *L. braziliensis*, the tested structures had similar VINA score values, except for derivative **302a**, which presented the lowest affinity value (-11.17 kcal/mol). All structures had lower docking values compared to MTX (-9.64 kcal/mol), as seen in Table 3.

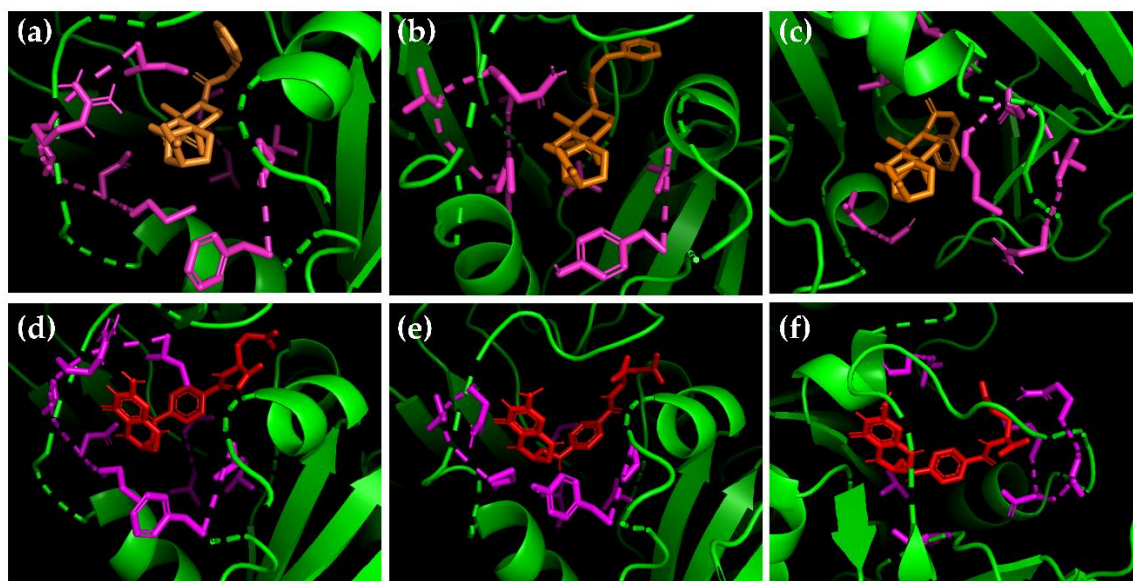


Figure 3. Best pose of structure **302** (orange) and MTX (red) in the active site of (a-d) *L. amazonensis* (b-e) *L. braziliensis* and (c-f) *L. panamensis* DHFR-TS (green). Flexible aminoacids are marked in pink.

By analyzing the interactions between the tested kauranes and the flexible residues of the active site of *L. braziliensis* DHFR-TS, it was found that the unsaturation of carbon-17 is crucial for the inhibition of **301**, **302**, and their derivatives **301a** and **302a** with the enzyme, via π -alkyl interactions with Y91 and M53. This interaction was also observed in MTX through a π -sulfur interaction with the thiol group of methionine. In addition, the potential inhibitory activity observed for structure **4** was related to a hydrogen bond interaction between Q48 and the carbonyl group of carbon-3, as well as the presence of a 1,3-dioxolane group. Neither structure **301** nor **302** interacted with the phenylpropanoid moiety of their structures, which was different from what was previously observed with *L. major* DHFR-TS.

For *L. panamensis*, both derivative structures **301a** and **302a** presented the lowest VINA score values, -12.55 kcal/mol and -12.54 kcal/mol, respectively, showing a higher inhibitory activity compared to the four kauranes and the control, MTX (Table 3). Structures **301** and **302** did not show any π -alkyl interaction with Y91 (Figure 3b), with mainly Van der Waals forces observed with flexible residues such as V31, V49, and V156. Kaurane **301** established a hydrogen bond between the carboxylic group of Carbon 4 and residue Q48. This interaction was also observed for MTX, however, a Negative-Negative unfavorable interaction with D52 affected the affinity value for this compound. A common

alkyl interaction between V31 and the unsaturation of carbon-17 of structure **135**, and between V31 and the 1,3-dioxolane group of structure **4**, was also observed.

In the same manner, *L. amazonensis* exhibited a behavior similar to that of *L. braziliensis* with VINA score values ranging from -10.52 to -11.14 kcal/mol, all of which showed lower affinity values compared to MTX (-9.54 kcal/mol). The latter only showed three interactions with the flexible residues in the active site of the enzyme, including two Van der Waals interactions with V49 and Q48 and a π -sulfur interaction between the sulfhydryl group of M53 and the pteridine ring. Structures **301** and **302** displayed the same interactions, which were classified into three groups: π -alkyl with M53 and Y91, alkyl with V87, and Van der Waals with Q49, V31, and V156. On the other hand, Structure **4** was the only kaurane that exhibited a hydrogen bond interaction with Q48, which might explain the slight difference in its affinity value.

Figure 3 displays the complex between the best-docked pose of structure **302**, the potential multispecies dual DHFR-TS/PTR1 inhibitor, and each of the three DHFR-TS hybrid models built in this study. For the *L. braziliensis* and *L. amazonensis* species (Figure 3a and 3b), similar poses and intermolecular interactions were observed, highlighting, the π -alkyl interactions of Y91 and M53 with the double bond of carbon-17.

In contrast, *L. panamensis* showed a different three-dimensional conformation in the active site of DHFR-TS, with a different spatial position for the phenylpropanoid moiety compared to the other two species of *Leishmania*. Additionally, residue Y91, which was a key residue in the interaction of the evaluated structures with the enzyme in *L. amazonensis* and *L. braziliensis* species (Figure 3a and 3b), did not interact with the unsaturation of carbon-17. This same pattern was also observed for MTX, where Y91 did not appear to be a relevant amino acid for the inhibitory activity.

2.4. Molecular dynamics simulations for *L. major* and *L. braziliensis* DHFR-TS interacting with **302 and MTX.**

To validate the hybrid models built for the different *Leishmania* species used in this study and evaluate the protein-ligand stability of structure **302** and its derivative **302a**, molecular dynamics (MD) studies were performed on *L. major* and *L. braziliensis* DHFR-TS using MTX as a reference ligand.

Initially, Root-mean-square deviation (RMSD) analyses were conducted to assess the structural stability of the receptor frame. These analyses measured the distance between different positions of a set of atoms over time (in nm) [30]. For *L. major* DHFR-TS, during

the first 30 ns, similar levels of perturbation were observed, with RMSD values ranging from 0.15 to 0.30 nm for structures **302**, **302a**, MTX, and the apoenzyme (apoLmDHFR-TS, the protein without the ligand).

After 30 ns, the protein in complex with structure **302** and its derivative **302a** showed increased stability, with lower RMSD values compared to apoLmDHFR-TS (Figure 4a). The same pattern was observed for *L. braziliensis*, as apoLbDHFR-TS showed a constant increase in RMSD values (from 0.20 to 0.40 nm) during the 50 ns, while the complexes LbDHFR-TS:302 and LbDHFR-TS:MTX had values ranging from 0.20 to 0.25 nm (Figure 4b). This indicates that structure 302 increases the stability of the complex with DHFR-TS in both *Leishmania* species, similar to the stability exerted by MTX.

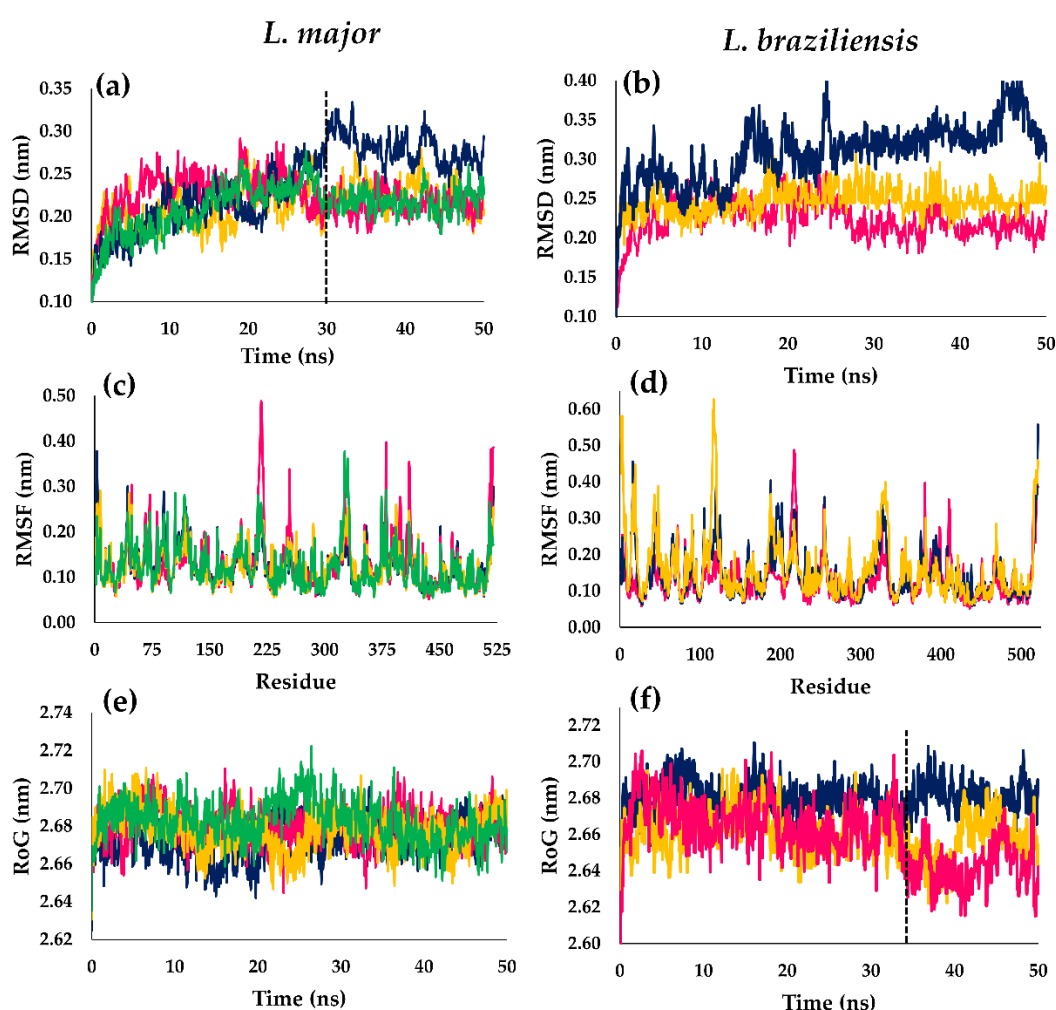


Figure 4. (a-b) Root-mean-square deviation (RMSD), (c-d) root-mean-square-fluctuation (RMSF), and (e-f) radius of gyration (RoG) values within the *L. major* DHFR-TS and *L. braziliensis* DHFR-TS binding site, obtained after molecular dynamics simulations. Apoenzyme (blue); DHFR-TS:MTX complex (yellow); DHFR-TS:**302** complex (pink); DHFR-TS:**302a** complex (light green).

Afterward, we analyzed the flexibility of residues with different ligands using root-mean-square fluctuations (RMSF) values. Similar patterns were found in both *L. major* and *L. braziliensis* during the entire dynamic simulations (Figure 4c and 4d). Regions with defined tertiary structures (α -helices or β -sheets) showed similar RMSF values (0.1 to 0.2 nm) for structure **302** and its derivative **302a** in complex with *L. major* DHFR-TS, as well as for the apoenzyme.

However, the control compound MTX presented higher RMSF values, particularly in loop regions of the protein. On analyzing the RMSF values in *L. braziliensis* DHFR-TS, the *Lb*DHFR-TS:**302** and the apoenzyme showed similar behaviors over the simulation time, while structure **302** had higher fluctuations in loop regions than MTX and the uncomplexed protein, especially in the region from A113 to T121, where values ranging from 0.30 nm to 0.63 nm were observed. Despite this, in regions with defined tertiary structure, both MTX and the kaurane **302** showed RMSF values lower than 0.20 nm, which indicates low flexibility in *L. major* DHFR-TS when complexed (Figure 4d).

In addition, we observed the evolution of the packing level of *L. major* and *L. braziliensis* DHFR-TS through the radius of gyration (RoG) values. For *L. major*, the complexes with structure **302** and its derivative **302a** showed no difference in RoG values compared with the control MTX and apo*Lm*DHFR-TS (ranging from 2.65 nm to 2.70 nm), indicating high stability and low fluctuations in the tertiary structure (Figure 4e).

For *L. braziliensis*, the RoG values for DHFR-TS were different for the two evaluated complexes compared to the apo*Lb*DHFR-TS. During the first 30 ns of the simulation, no differences in RoG values were observed (RoG of approximately 2.68 nm). However, after this time, the complexes *Lb*DHFR-TS:**302** and *Lb*DHFR-TS:MTX demonstrated different behaviors, with a reduction in the RoG value (approximately 2.64 nm). This indicates that structure **302** stably folded after the simulation, compared to the apoenzyme, which remained at a constant value during the 50 ns test period (Figure 4f).

2.5 Free energy calculations by the Molecular Mechanics - Poisson Boltzmann Surface Area approach (MM/PBSA) method.

After the molecular dynamic simulations were completed, the binding free energies for the complexes of structures **302** and **302a**, and MTX with *L. major* DHFR-TS, as well as the complexes of structure **302** and MTX with *L. braziliensis* DHFR-TS, were calculated using the MM/PBSA method. The kaurane **302** and its derivative **302a** in complex with *L. major* DHFR-TS reached binding free energy values of -105.8 kJ/mol and -118.2 kJ/mol,

respectively, which were both higher than the value measured for the complex *Lm*DHFR-TS: MTX, which was -127.4 kJ/mol. Conversely, the complex *Lb*DHFR-TS:**302** (-137.2 kJ/mol) reached a lower binding free energy value compared to the complex *Lb*DHFR-TS: MTX (-93.9 kJ/mol). Nevertheless, for both *Leishmania* species, similar energetic contributions were observed, which were linked to the structural features of the evaluated molecules (Table 4).

Table 4. Binding free energies (kJ/mol) from the MM/PBSA calculations for structure **302** and its derivative **302a** for *L. major* DHFR-TS and Structure **302** for *L. braziliensis* DHFR-TS; In both proteins MTX was used as reference ligand.

<i>Leishmania major</i>					
Structure	Van der Waals (kJ/mol)	Electrostatic (kJ/mol)	Polar solvation (kJ/mol)	SASA (kJ/mol)	Binding energy (kJ/mol)
302	-211.1 ± 6.9	-26.4 ± 3.2	151.6 ± 8.1	-19.8 ± 1.6	-105.8 ± 5.0
302a	-219.2 ± 5.2	-27.6 ± 3.0	148.5 ± 8.6	-19.9 ± 1.0	-118.2 ± 4.5
MTX	-233.9 ± 7.4	-56.3 ± 5.2	192.4 ± 9.5	-22.1 ± 2.2	-127.4 ± 6.1
<i>Leishmania braziliensis</i>					
302	-230.1 ± 6.0	-20.6 ± 1.5	137.2 ± 7.5	-24.0 ± 0.2	-137.2 ± 6.6
MTX	-206.4 ± 5.4	-54.8 ± 4.5	189.6 ± 10.0	-21.3 ± 0.4	-93.3 ± 4.9

For the complexes with **302** in both *Leishmania* species, Van der Waals, Electrostatic, and solvent-accessible surface area (SASA) parameters showed negative contributions to the binding free energy. The Van der Waals parameter had the highest negative contribution, and these results are directly related to the molecular docking calculations, where in *L. major* and mainly in new world *Leishmania* species, this type of interaction is fundamental for the stability of the DHFR-TS-diterpenoid complexes. Electrostatic parameter also contributed negatively to the binding free energies; however, its contribution was close to 50% for structure **302** and its derivative **302a**, compared to the contribution observed for MTX, which had a higher contribution to the total binding energy. Finally, for all molecules, polar solvation had a positive contribution to the total energy value, with larger contributions to the complexes with MTX in both evaluated *Leishmania* species.

3. Materials and Methods

3.1. *LmDHFR*-TS enzyme inhibition assay

Purification and kinetic characterization of the recombinant *LmDHFR*-TS protein was performed according to the previously reported procedures [31,32]. The *in vitro* evaluation of selected diterpenoids (i.e., **4**, **135**, **301**, **302**, **301a**, and **302a**) for inhibitory activity against *LmDHFR*-TS was conducted using a spectrophotometric assay under standard DHFR conditions. The assay consisted of *LmDHFR* (2.7 nM), bovine serum albumin (BSA, 1 mg/mL), N-[tris(hydroxymethyl)-methyl]-2-aminoethanesulfonic acid (TES) buffer (100 mM, pH 7.0, 150 mM β -mercaptoethanol, 2 mM ethylenediaminetetraacetic acid (EDTA)), and nicotinamide adenine dinucleotide phosphate (NADPH, 100 μ M) with varying concentrations of the test compounds (0.1-128 μ M).

The reaction was initiated by adding the substrate (7,8-dihydrofolate (H₂F), 20 μ M) and was monitored for 360 seconds at 340 nm (i.e. oxidation of NADPH to NADP⁺) to determine the initial reaction rate (V_o) through linear regression analysis of the resulting absorbance profile. All measurements were performed in triplicate and MTX was used as a positive control. The resulting V_o values were used to calculate the % inhibition, as $100 - (R_i / R_c \times 100)$, where R_i is the V_o in the presence of the inhibitor and R_c is the V_o in the absence of inhibitors (1% DMSO v/v final concentration).

The % inhibition was measured for at least five concentrations (0.1-128 μ M) for each test compound (diterpenoids and MTX), and concentration-response curves (% inhibition vs. Log[inhibitor]) were constructed using non-linear regression to determine the IC₅₀ using GraphPad Prism 7.0 (GraphPad, San Diego, CA, USA). The K_{iapp} values were finally calculated using the Cheng-Prusoff equation for competitive inhibition with a 1:1 stoichiometry and reversible inhibitor-binding reactions: $K_{iapp} = IC_{50} / (1 + [S] / K_m)$, where [S] is the substrate (H₂F) concentration and K_m is the Michaelis constant. The substrate K_m was calculated during the kinetic characterization of the purified, recombinant *LmDHFR*-TS and was determined to be $2.4 \pm 0.7 \mu$ M.

3.2. Isolation of compound **148**

Kaurane-type diterpene **148** was isolated from *Euphorbia graminea* Jacq. (Euphorbiaceae), which was propagated under greenhouse conditions from commercially available seeds (Swallowtail Garden Seeds, Santa Rosa, CA, USA). The aerial part (128 g)

of two-month-old plants of *E. graminea* was extracted with 96% ethanol, and the raw extract (11.2 g) was purified by column chromatography (CC) using a gradient elution of n-hexane to methanol, yielding fifteen fractions. The purification of fraction 7 was then performed independently by flash column chromatography on SiO₂ using a mobile phase of a 7:3 mixture of n-hexane and ethyl acetate, which resulted in the isolation of diterpene 148 (35.6 mg). Its spectroscopic data, including NMR and HRMS, were found to match those of the previously isolated compound ent-kaurane-3-oxo-16 α ,17-diol [33]

3.3. Synthesis of 16 β ,17-isopropylidenedioxy-ent-kauran-3-one (4)

Compound **4** was synthesized from **148** using a previously reported procedure [34]. Briefly, compound **148** (24 mg, 0.075 mmol) and tetrahydrofuran (THF) (4 mL) were mixed in a 10-mL round-bottom flask by stirring at 0 °C. Then, 2,2-dimethoxypropane (46 μ L, 0.375 mmol) and *p*-toluenesulfonic acid monohydrate (0.75 mg, 0.375 μ mol) were added. The reaction mixture was stirred at 0 °C for 2 hours, allowed to warm to 20 °C, and then stirred at this temperature for 16 hours.

The reaction was then quenched with saturated NaHCO₃ (3 mL) and extracted with CH₂Cl₂ (3 x 3 mL). The CH₂Cl₂ extract was separated, washed with 10% NaCl (2 x 3 mL), dried over MgSO₄, filtered, and concentrated under reduced pressure to obtain the structure **4** (26 mg, 96%); [α]_D²⁰ –41.8 (c 0.04, CHCl₃); ¹H NMR (400 MHz, CDCl₃) δ _H 4.22 (d, J = 8.3 Hz, 1H), 3.61 (d, J=10.5 Hz, 1H), 2.41 (dd, J = 8.1, 6.3 Hz, 2H), 2.33-2.28 (m, 1H), 1.85 (dd, J = 10.4, 3.6 Hz, 1H), 1.76-1.72 (m, 1H), 1.68-1.65 (m, 2H), 1.64-1.61 (m, 1H), 1.52-1.47 (m, 1H), 1.47-1.44 (m, 1H), 1.37 (s, 3H), 1.33 (s, 3H), 1.27-1.25 (m, 3H), 1.24-1.23 (m, 1H), 1.23-1.19 (m, 3H), 1.14-1.11 (m, 1H), 0.99 (s, 3H), 0.91 (s, 3H), 0.88 (s, 3H), 0.82 (d, J = 8.4 Hz, 1H); ¹³C NMR (100 MHz, CHCl₃) δ _C 217.6, 193.0, 109.3, 79.6, 69.5, 55.2, 54.7, 52.3, 47.4, 44.1, 40.6, 40.4, 37.8, 37.3, 37.2, 34.5, 27.6, 27.4, 27.2, 26.7, 20.3, 19.4, 17.8; HREIMS [M+H]⁺ *m/z* 361.2724 (calcd. for C₂₃H₃₇O₃, 361.2743).

3.4. Hybrid models of *Leishmania* DHFR-TS

Hybrid models of the dihydrofolate reductase-thymidylate synthase (DHFR-TS) of different *Leishmania* species were constructed using YASARA software (YASARA 18.4.24, Vienna, Austria: YASARA Biosciences GmbH, 2018). The FASTA sequences of *L. major* DHFR-TS (P07382), *L. braziliensis* DHFR-TS (A4H4P8), *L. panamensis* DHFR-TS (S5M3K7), and *L. amazonensis* DHFR-TS (P16126) were obtained from the UNIPROT database (<https://www.uniprot.org/>). The constructed hybrid models were validated

through stereochemical quality assessment using PROCHECK [35]. PROCHECK evaluated molecular diversity through several stereochemical parameters, including the torsional angles of the main chain, side chain torsional angles, bad contacts or steric impediments, and planarity.

PROCHECK generated a Ramachandran graph [23], which verified the allowed and prohibited regions of the main amino acid chain. The structural quality was evaluated using VERIFY 3D software (<https://services.mbi.ucla.edu/SAVES/>, accessed on January 3, 2023) and WHAT IF (<https://swift.cmbi.ru.nl/servers/html/index.html>, accessed on January 5, 2023). VERIFY 3D software analyzes the compatibility of the protein sequence with its 3D structure based on the chemical environment, while WHAT IF analyzes various structural parameters, such as atomic contacts between residues. The Discovery Studio Visualizer (BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1.0.20298, San Diego: Dassault Systèmes, 2020) was used to visualize the modeled protein [20].

3.5. Molecular docking calculations

The hybrid model of *L. major* DHFR-TS in complex with methotrexate (PDB ID: MTX) was used for the molecular docking calculations of the six kaurane-type diterpenes using the Molegro 6.0.1 software. All water molecules were removed from the enzyme structures and both the enzyme and compound structures were prepared with the same default parameters in the same software package. The MolDock Score function was used as the score function, and the internal ES, internal H-bond, Sp2-Sp2 Torsions were all checked as the ligand evaluation criteria. The molecular docking procedure was run 10 times, using the MolDock SE algorithm, with a maximum of 1500 interactions, a maximum population size of 50, a maximum of 300 steps, a neighbor distance factor of 1.00, and a maximum of 5 poses returned. A grid with a 15 Å radius and 0.30 Å resolution was used to cover the ligand-binding site for the enzyme structure [36, 37].

For *L. brasiliensis*, *L. panamensis*, and *L. amazonensis* DHFR-TS, the docking calculations were performed using the Autodock/Vina (1.1.2) plug-in for PyMOL (1.3r2) under a Python 2.5.2 environment for Windows. The minimized structure was located in a cube with dimensions of 22.5 Å × 22.5 Å × 22.5 Å and a grid spacing of 0.375 Å at the geometric center of the binding pocket (coordinates for *L. brasiliensis*: 43.01, 23.70, 1.67; *L. panamensis*: 43.64, 24.00, 1.50; and *L. amazonensis*: 43.49, 24.96, 1.95), which was identified through cavities analysis in Molegro 6.0.1. Flexible residues in the binding site were selected for each model: *L. brasiliensis* and *L. amazonensis*: I30, V31, Q48, V49,

M53, V87, and V156; *L. panamensis*: I30, V31, Q48, V49, D52, M53, S86, and V87. The docking poses were classified based on their docking scores, such as the free energy or affinity, and each calculation was performed in three replicates. Methotrexate (MTX) was used as a control. The two-dimensional residual interaction diagrams were visualized on the Discovery Studio Visualizer (BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1.0.20298, San Diego: Dassault Systèmes, 2020) [20].

3.6. Molecular dynamics simulations

Molecular dynamics simulations were carried out using Gromacs 5.0.5 on an Ubuntu 12.04 server [38,39]. The structures **302**, its derivative **302a**, MTX, and the hybrid models of *L. brasiliensis*, *L. panamensis*, and *L. amazonensis* DHFR-TS were used as inputs for the simulations. [40] The hydrogen atoms and corresponding charges for the ligands were added using the AM1-BCC charge scheme in UCSF Chimera, and the ligand topologies were generated automatically with the ACPYPE script.

The protein topologies were obtained in Gromacs using the Amber 99SB force field and the TIP3P water model. Solvation was performed in a triclinic box with a margin distance of 1.0 nm and 0.1 M NaCl was added to the complexes and proteins by randomly replacing water molecules until neutrality was achieved [30, 37-39].

The systems were energy-minimized for 2,000 steps using the steepest descent method. Then, NVT equilibration was performed at 310 K for 50 ps followed by NPT equilibration for 500 ps, using the Parrinello-Rahman method at 1 bar with position restraints. The solute position restraints were then released, and a production run was performed for 5 ns while maintaining constant temperature and pressure at 310 K and 1 bar, respectively.

The coordinates were recorded in a 1 fs time step, and electrostatic forces were calculated using the particle-mesh Ewald method. All simulations used periodic boundary conditions and covalent bond lengths were constrained by the LINCS algorithm. The binding free energies were calculated using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method based on the trajectories obtained from the molecular dynamics' simulations [37-39].

4. Conclusions

This study identified compounds **302** (3 α -cinnamoyloxy-ent-kaur-16-en-19-oic acid) and **302a** as potential inhibitors of both PTR1 and DHFR-TS in *L. major*, building

upon previous findings of PTR1 inhibition [20]. Both **302** and **302a** displayed *in vitro* inhibitory activity against *L. major* DHFR-TS, with IC₅₀ values of 6.3 and 4.5 μ M, respectively. Additionally, other kaurane-type diterpenes, such as synthesized structure **4**, also inhibited DHFR-TS *in vitro* with an IC₅₀ value of 7.6 μ M. Structures **301** and **301a**, which were previously classified as inactive against PTR1, also showed inhibitory activity against DHFR-TS, verifying the results obtained from MolpredictX. Furthermore, molecular docking calculations using a hybrid model of *L. major* DHFR-TS allowed evaluation of the mechanism of action of the tested kauranes. The *p*-hydroxyl group of the phenylpropanoid moiety of structure **302a** was found to play a crucial role in the inhibition of DHFR-TS.

Additionally, hybrid models for three *Leishmania* species with high incidence in Central and South America were constructed. The best docked results for structure **302** and its derivative **302a** in the three hybrid models showed a correlation between the affinity values obtained from the molecular docking and some structural features of the kauranes, such as the presence of an unsaturation at carbon-17 that interacts with the amino acids of DHFR-TS through π -alkyl interactions, making these two structures potential multispecies inhibitors. Furthermore, molecular dynamics' simulation, in addition to validating the hybrid models, confirmed the results previously obtained from the molecular docking calculations. So, this study presented a valuable approach for identifying potential dual PTR1/DHFR-TS inhibitors, contributing to the development of alternative chemotherapy strategies against these diseases.

References

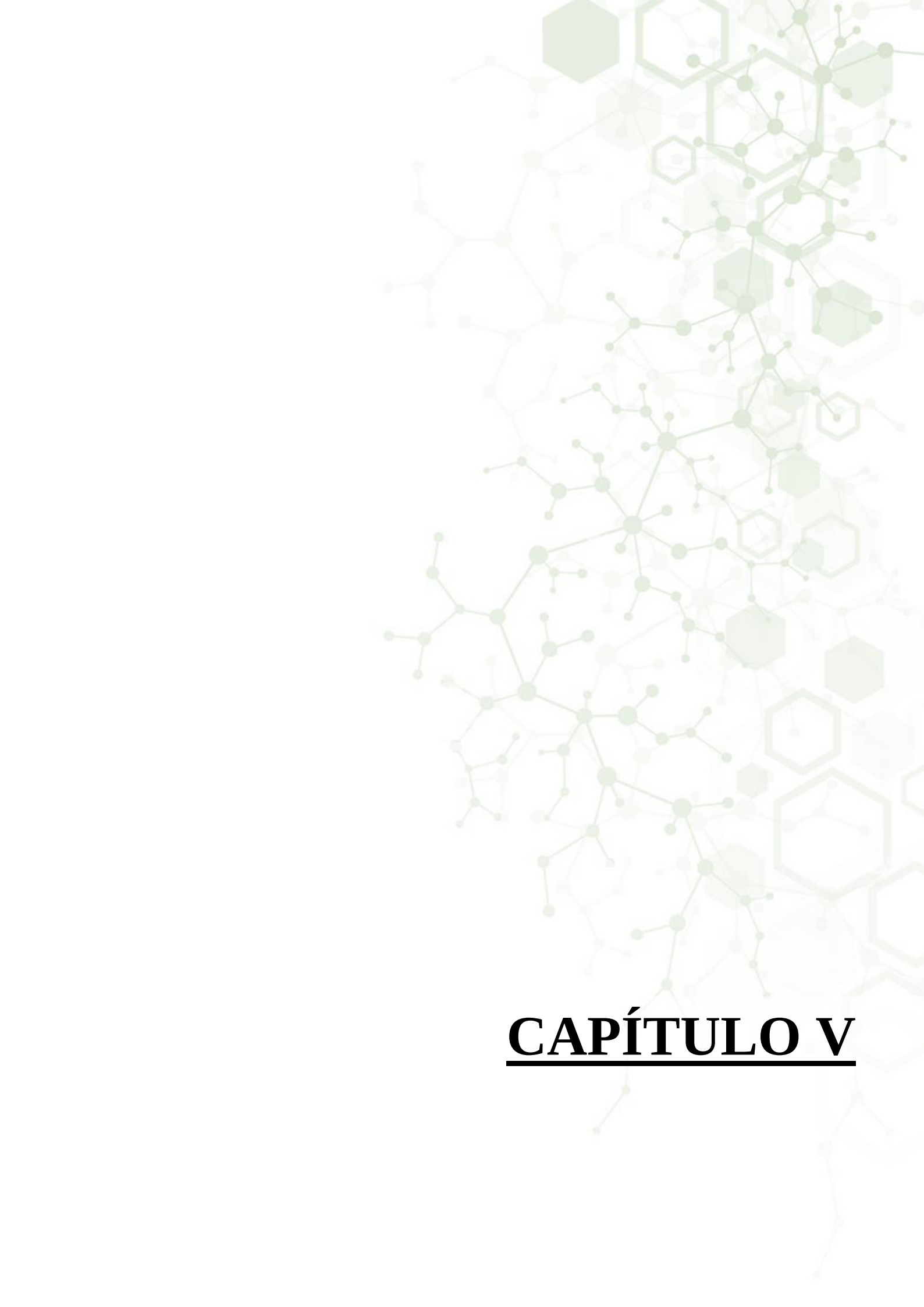
- [1]. Soni, M.; Pratap, J.V. Development of Novel Anti-Leishmanials: The Case for Structure-Based Approaches. *Pathogens* **2022**, *11*, 950.
- [2]. Gouri, V.; Upreti, S.; Samant, M. Evaluation of target-specific natural compounds for drug discovery against leishmaniasis. *Parasitol. Int.* **2022**, *91*, 102622.
- [3]. Salari, S.; Bamorovat, M.; Sharifi, I.; Almani, P.G.N. Global distribution of treatment resistance gene markers for leishmaniasis. *J. Clin. Lab. Anal.* **2022**, *36*, e24599.
- [4]. Sasidharan, S.; Saudagar, P. Leishmaniasis: where are we and where are we heading? *Parasitol. Res.* **2021**, *120*, 1541–1554.
- [5]. Gupta, D.; Singh, P.K.; Yadav, P.K.; Narender, T.; Patil, U.K.; Jain, S.K.; Chourasia, M.K. Emerging strategies and challenges of molecular therapeutics in antileishmanial drug development. *Int. Immunopharmacol.* **2023**, *115*, 109649.

- [6]. Uliana, S.R.B.; Trinconi, C.T.; Coelho, A.C. Chemotherapy of leishmaniasis: Present challenges. *Parasitology* **2018**, *145*, 464–480.
- [7]. Brindha, J.; Balamurali, M.M.; Chanda, K. An Overview on the Therapeutics of Neglected Infectious Diseases—Leishmaniasis and Chagas Diseases. *Front. Chem.* **2021**, *9*, 622286.
- [8]. Vickers, T.J.; Beverley, S.M. Folate metabolic pathways in Leishmania. *Essays Biochem.* **2011**, *51*, 63–80.
- [9]. Gilbert, I.H. Inhibitors of dihydrofolate reductase in leishmania and trypanosomes. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2002**, *1587*, 249–257.
- [10]. Ivanetich, K.M.; Santi, D. V Thymidylate synthase-dihydrofolate reductase in protozoa. *Exp. Parasitol.* **1990**, *70*, 367–371.
- [11]. Shamshad, H.; Bakri, R.; Mirza, A.Z. Dihydrofolate reductase, thymidylate synthase, and serine hydroxy methyltransferase: successful targets against some infectious diseases. *Mol. Biol. Rep.* **2022**, *49*, 6659–6691.
- [12]. Ong, H.B.; Sienkiewicz, N.; Wyllie, S.; Fairlamb, A.H. Dissecting the metabolic roles of pteridine reductase 1 in Trypanosoma brucei and Leishmania major. *J. Biol. Chem.* **2011**, *286*, 10429–10438.
- [13]. Nare, B.; Luba, J.; Hardy, L.W.; Beverley, S. New approaches to Leishmania chemotherapy: Pteridine Reductase 1 (PTR1) as a target and modulator of antifolate sensitivity. *Parasitology* **1997**, *114*, S101-10.
- [14]. Panecka-Hofman, J.; Poehner, I.; Wade, R.C. Anti-trypanosomatid structure-based drug design - lessons learned from targeting the folate pathway. *Expert Opin. Drug Discov.* **2022**, *17*, 1029–1045.
- [15]. Kheirandish, F.; Bandehpour, M.; Haghighi, A.; Mahboudi, F.; Mohebbali, M.; Kazemi, B. Inhibition of Leishmania major PTR1 Gene Expression by Antisense in Escherichia coli. *Iran. J. Public Health* **2012**, *41*, 65–71.
- [16]. Das Neves, G.M.H.; Kagami, L.P.; Gonçalves, I.L.; Eifler-Lima, V.L. Targeting pteridine reductase 1 and dihydrofolate reductase: The old is a new trend for leishmaniasis drug discovery. *Future Med. Chem.* **2019**, *11*, 207–2130.
- [17]. Nogueira, M.S.; Da Costa, F.B.; Brun, R.; Kaiser, M.; Schmidt, T.J. Ent-pimarane and ent-kaurane diterpenes from Aldama discolor (Asteraceae) and their antiprotozoal activity. *Molecules* **2016**, *21*, 1237.
- [18]. Miranda, M.M.; Panis, C.; Da Silva, S.S.; Macri, J.A.; Kawakami, N.Y.; Hayashida, T.H.; Madeira, T.B.; Acquaro, V.R.; Nixdorf, S.L.; Pizzatti, L.; et al. Kaurenoic acid

possesses leishmanicidal activity by triggering a NLRP12/IL-1 β /cNOS/NO Pathway. *Mediators Inflamm.* **2015**, 2015, 392918.

- [19]. dos Santos, A.O.; Izumi, E.; Ueda-Nakamura, T.; Dias-Filho, B.P.; da Veiga-Júnior, V.F.; Vataru Nakamura, C. Antileishmanial activity of diterpene acids in copaiba oil. *Mem. Inst. Oswaldo Cruz* **2013**, 108, 59–64.
- [20]. Herrera-Acevedo, C.; Flores-Gaspar, A.; Scotti, L.; Mendonça-Junior, F. J. B.; Scotti, M. T.; Coy-Barrera, E. Identification of Kaurane-type diterpenes as inhibitors of Leishmania pteridine reductase I. *Molecules* **2021**, 26(11), 3076.
- [21]. Cheng, Y.C.; Prusoff, W.H. the concentration of inhibitor which causes 50 percent inhibition (I) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, 22, 3099–3108.
- [22]. Tullius Scotti, M.; Herrera - Acevedo, C.; Barros de Menezes, R. P.; Martin, H. J.; Muratov, E. N.; Ítalo de Souza Silva, Á. Et al. MolPredictX: Online Biological Activity Predictions by Machine Learning Models. *Mol. Inform.* **2022**, 2200133.
- [23]. Lovell, S.C.; Davis, I.W.; Arendall Iii, W.B.; De Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by C α geometry: ϕ , ψ , and C β deviation. *Proteins* **2003**, 50, 437–450.
- [24]. Vriend, G.; Sander, C. Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.* **1993**, 26(1), 47-60.
- [25]. Kosaka, A.; Sakamoto, N.; Hikone, M.; Imai, K.; Ota, M.; Washino, T.; Iwabuchi, S. (2020). Failure of liposomal-amphotericin B treatment for new world cutaneous leishmaniasis due to leishmania braziliensis. *Intern. Med.* **2020**, 59(9), 1227-1230.
- [26]. Herrera, G.; Barragán, N.; Luna, N.; Martínez, D.; De Martino, F.; Medina, J; et al. An interactive database of Leishmania species distribution in the Americas. *Sci. data* **2020**, 7(1) 110.
- [27]. Davila, M.; Pineda, V.; Calzada, J. E.; Saldaña, A.; Samudio, F. Evaluation of cytochrome b sequence to identify Leishmania species and variants: the case of Panama. *Memórias do Instituto Oswaldo Cruz*, **2021**, 116.
- [28]. Rodrigues, M. P.; Tomaz, D. C.; de Souza, L. A.; Onofre, T. S.; de Menezes, W. A.; Almeida-Silva, J.; et al. Synthesis of cinnamic acid derivatives and leishmanicidal activity against Leishmania braziliensis. *Eur. J. Med. Chem.* **2019**, 183, 111688.
- [29]. Brustolin, A. Á.; Ramos-Milaré, Á. C. F. H.; de Mello, T. F. P.; Aristides, S. M. A.; Lonardoní, M. V. C.; Silveira, T. G. V. In vitro activity of cinnamaldehyde on Leishmania (Leishmania) amazonensis. *Exp. Parasitol.* **2022**, 236, 108244.

- [30]. Rozo-Lugo, C.; Cuca-Suárez, L.E.; Schmidt, T.J.; Coy-Barrera, E. Tetrahydrobenzofuran-6 (2 H)-one neolignans from *Ocotea heterochroma*: Their platelet activating factor (PAF) antagonistic activity and in silico insights into the PAF receptor binding mode. *J. Nat. Prod.* **2018**, *81*, 1968–1975
- [31]. Grumont, R.; Sirawaraporn, W.; Santi, D. V Heterologous Expression of the Bifunctional Thymidylate Synthase-Dihydrofolate Reductase from *Leishmania major*. *Biochemistry* **1988**, *27*, 3776–3784.
- [32]. Nare, B.; Hardy, L.W.; Beverley, S.M. The roles of pteridine reductase 1 and dihydrofolate reductase- thymidylate synthase in pteridine metabolism in the protozoan parasite *Leishmania major*. *J. Biol. Chem.* **1997**, *272*, 13883–13891.
- [33]. Yi-Li, D.; Zhong-Jian, J. Tetracyclic diterpenols from *Euphorbia sieboldiana*. *Phytochemistry* **1991**, *30*, 2413–2415.
- [34]. Bon, D.J.-Y.D.; Banwell, M.G.; Willis, A.C. A chemoenzymatic total synthesis of the hirsutene-type sesquiterpene (+)-connatusin B from toluene. *Tetrahedron* **2010**, *66*, 7807–7814.
- [35]. 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
- [36]. Acevedo, C.H.; Scotti, L.; Scotti, M.T. In silico studies designed to select sesquiterpene lactones with potential antichagasic activity from an in-house asteraceae database. *ChemMedChem* **2018**, *13*, 634–645.
- [37]. Herrera-Acevedo, C.; Maia, M.D.S.; Cavalcanti, É.B.V.S.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Selection of antileishmanial sesquiterpene lactones from Sistemax database using a combined ligand-/structure-based virtual screening approach. *Mol. Divers.* **2020**, 1–17
- [38]. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- [39]. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- [40]. Krieger, E.; Vriend, G. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, **2014**, *30*, 2981–2982.



CAPÍTULO V

O ácido cinâmico é um ácido orgânico presente naturalmente em plantas, sendo caracterizado por baixa toxicidade e amplo espectro de atividades biológicas. Na busca por compostos farmacologicamente ativos, os derivados de ácido cinâmico são considerados importantes e promissores, com grande potencial para o desenvolvimento de medicamentos [1]. Quimicamente, nos ácidos cinâmicos, a funcionalidade do ácido 3-fenil acrílico oferece três principais sítios reativos: substituição no anel fenílico, adição na insaturação- α , β e reações da funcionalidade do ácido carboxílico [2]. Derivados de ácido cinâmico, incluindo flavonoides e lignanas, apresentam notável potencial leishmanicida, tornando-os candidatos promissores no combate a infecções por *Leishmania*.

Encontrado em muitas plantas, o ácido cinâmico e seus derivados são frequentemente avaliados por sua atividade farmacológica [3]. Gouri et al. relatam alguns inibidores naturais contra amastigotas de *Leishmania*, como luteolina ($IC_{50} = 3,12 \mu M$), quercetina ($IC_{50} = 10,5 \mu M$), crisina ($IC_{50} = 13 \mu M$), apigenina, miricetina, ácido cinâmico ($IC_{50} = 0,25 \mu M$) e licocalcona A ($IC_{50} = 0,9 \mu M$), que podem desempenhar um papel importante na descoberta de medicamentos [4]. Por outro lado, Peixoto et al. avaliaram a atividade biológica de 25 derivados de ácido cinâmico contra amastigotas de *Leishmania braziliensis*, obtendo resultados promissores e observando que anéis aromáticos com oxigênio como heteroátomo tiveram um efeito benéfico em termos de atividade contra *Leishmania* [5].

Flavonoides derivados do ácido cinâmico têm mostrado notável atividade leishmanicida. Compostos como quercetina [6], kaempferol [7] e naringenina [8], originados da via do ácido cinâmico, possuem propriedades anti-*Leishmania*. Sua capacidade de modular o estresse oxidativo e interferir em processos celulares essenciais no parasita os torna valiosos no combate a infecções por *Leishmania*.

Lignanas, outra classe de derivados do ácido cinâmico, exibem atividades leishmanicidas promissoras. A nirantina, uma lignana isolada das partes aéreas da planta *Phyllanthus amarus*, apresenta amplo espectro de atividades farmacológicas, incluindo um potente efeito antileishmanial [9]. Secoisolariciresinol [10] e matairesinol [11], derivados do ácido cinâmico, demonstraram eficácia contra parasitas da *Leishmania*. Essas lignanas podem interferir nos mecanismos de crescimento e sobrevivência do parasita, tornando-as candidatas potenciais para o desenvolvimento de medicamentos.

Este capítulo foca na identificação de potenciais inibidores da diidrofolato redutase-timidilato sintase (DHFR-TS) da *Leishmania major*, essencial para a síntese de DNA. Utilizando um banco de dados com 314 metabólitos secundários derivados do ácido

cinâmico na família Asteraceae, uma triagem virtual identificou hits promissores, incluindo ácido litospérmico, diarctigenina e isolappaol A. Simulações de dinâmica molecular validaram sua estabilidade, e ensaios *in vitro* demonstraram inibição efetiva da *LmDHFR-TS*. Notavelmente, duas lignanas exibiram seletividade superior em relação ao metotrexato, indicando seu potencial no combate à leishmaniose. Pesquisas contínuas nessas lignanas híbridas butirolactonas C6C3 podem oferecer uma perspectiva mais promissora para combater essa doença tropical negligenciada.

REFERENCIAS

- [1].SOVA, Matej. Antioxidant and antimicrobial activities of cinnamic acid derivatives. **Mini reviews in medicinal chemistry**, v. 12, n. 8, p. 749-767, 2012.
- [2].DE, Prithwiraj; BALTAS, Michel; BEDOS-BELVAL, Florence. Cinnamic acid derivatives as anticancer agents-a review. **Current medicinal chemistry**, v. 18, n. 11, p. 1672-1703, 2011.
- [3].DE MORAIS, Mayara Castro et al. Antileishmanial Activity of Cinnamic Acid Derivatives against *Leishmania infantum*. **Molecules**, v. 28, n. 6, p. 2844, 2023.
- [4].GOURI, Vinita; UPRETI, Shobha; SAMANT, Mukesh. Evaluation of target-specific natural compounds for drug discovery against leishmaniasis. **Parasitology International**, p. 102622, 2022.
- [5].RODRIGUES, Michelle Peixoto et al. Synthesis of cinnamic acid derivatives and leishmanicidal activity against *Leishmania braziliensis*. **European Journal of Medicinal Chemistry**, v. 183, p. 111688, 2019.
- [6].ORYAN, Ahmad et al. Anti-leishmanial, immunomodulatory and anti-oxidative activity of quercetin against cutaneous leishmaniasis caused by *Leishmania major*. **Asian Pacific Journal of Tropical Biomedicine**, v. 13, n. 1, p. 26, 2023.
- [7].CHANDRASEKAR, R. et al. Therapeutic efficacy of flavonoids and terpenoids an ongoing herbal therapy in the treatment of leishmaniasis. **Nat Prod Ind J**, v. 14, n. 2, p. 124, 2018.
- [8].EZZAT, Shahira M.; SALEM, Mohamed A.; ZAYED, Ahmed. Antileishmanial Potentials of Phytochemicals. **Neglected Tropical Diseases and Phytochemicals in Drug Discovery**, p. 183-198, 2021.

- [9].CHOWDHURY, Sayan et al. The lignan niranthin poisons *Leishmania donovani* topoisomerase IB and favours a Th1 immune response in mice. **EMBO Molecular Medicine**, v. 4, n. 10, p. 1126-1143, 2012.
- [10].POSPÍŠIL, Jiří; KONRÁDOVÁ, Daniela; STRNAD, Miroslav. Antileishmanial Activity of Lignans, Neolignans, and Other Plant Phenols. **Progress in the Chemistry of Organic Natural Products 115**, p. 115-176, 2021.
- [11].TEPONNO, Rémy Bertrand; KUSARI, Souvik; SPITELLER, Michael. Recent advances in research on lignans and neolignans. **Natural product reports**, v. 33, n. 9, p. 1044-1092, 2016.
- [12].GILBERT, Ian H. Inhibitors of dihydrofolate reductase in *Leishmania* and trypanosomes. **Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease**, v. 1587, n. 2-3, p. 249-257, 2002.

Machine-Learning- and Structure-Based Virtual Screening for Selecting Cinnamic Acid Derivatives as *Leishmania major* DHFR-TS Inhibitors

Maria Camila Muñoz-Vega^{1,2}, Sofía López-Hernández¹ Adrián Sierra-Chavarro¹, Marcus Tullius Scotti³, Luciana Scotti³, Ericsson Coy-Barrera⁴, **Chonny Herrera-Acevedo^{1,3*}**

¹ Department of Chemical Engineering, Universidad ECCI, Carrera 19# 49-20, 111311, Bogotá D.C., Colombia.

² Laboratorio de Investigación en Biocatálisis y Biotransformaciones (LIBB), Grupo de Investigación en Ingeniería de los Procesos Agroalimentarios y Biotecnológicos (GIPAB), Departamento de Química, Universidad del Valle, Cali 760042, Colombia

³ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil; chonny622@gmail.com; renatabarros@ltf.ufpb.br; nferreiradesousa.nfs@gmail.com; luciana.scotti@gmail.com and mtscotti@gmail.com.

⁴ Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia; ericsson.coy@unimilitar.edu.co

*Correspondence: cherreraa@ecc.edu.co

Artigo aceitado em Molecules (IF: 4.60)

DOI Number

Data: 06 December 2023

Abstract: The critical enzyme dihydrofolate reductase-thymidylate synthase in *Leishmania major* (*LmDHFR-TS*) serves a dual-purpose role and is essential for DNA synthesis, a cornerstone of the parasite's reproductive processes. Consequently, the development of inhibitors against *LmDHFR-TS* is crucial for the creation of novel anti-*Leishmania* chemotherapies. In this study, we employed an in-house database containing 314 secondary metabolites derived from cinnamic acid that occurred in the Asteraceae family. We conducted a combined ligand/structure-based virtual screening to identify potential inhibitors against *LmDHFR-TS*. Through consensus analysis of both approaches,

we identified three compounds, i.e., lithospermic acid (**237**), diartigenin (**306**), and isolappaol A (**308**), that exhibited a high probability of being inhibitors according to both approaches and were consequently classified as promising hits. Subsequently, we expanded the binding mode examination of these compounds within the active site of the test enzyme through molecular dynamics simulations, revealing a high degree of structural stability and minimal fluctuations in its tertiary structure. The *in silico* predictions were then validated through *in vitro* assays to examine the inhibitory capacity of the top-ranked naturally occurring compounds against *Lm*DHFR-TS recombinant protein. The test compounds effectively inhibited the enzyme with IC₅₀ values ranging from 6.1 to 10.1 μ M. In contrast, other common cinnamic acid derivatives (i.e., flavonoid glycosides) from the Asteraceae family, such as hesperidin, isovitexin 4'-O-glucoside, and rutin, exhibited low activity against this target. The selective index (SI) for all tested compounds was determined using *Hs*DHFR with moderate inhibitory effect. Among these hits, lignans **306** and **308** demonstrated the highest selectivity, displaying superior SI values compared to methotrexate, the reference inhibitor of DHFR-TS. Therefore, continued research into the anti-leishmanial potential of these C6C3-hybrid butyrolactone lignans may offer a brighter outlook for combating this neglected tropical disease.

Keywords: *Leishmania*; Asteraceae; DHFR-TS; lignans; flavonoids; natural products; machine learning

1. Introduction

Leishmaniasis is a neglected tropical disease (NTD) caused by protozoan parasites of the genus *Leishmania*, which are transmitted by the bite of infected sandflies. This disease affects millions of people worldwide, particularly in developing countries with poor health infrastructure. The primary clinical forms of the disease are visceral, cutaneous, and mucocutaneous. According to the World Health Organization (WHO), the global burden of leishmaniasis is estimated to be around 700,000 to 1 million new cases each year, with 90% of the cases occurring in just six countries: Afghanistan, Algeria, Brazil, Colombia, Iran, and Syria [1,2]. The sandflies that transmit leishmaniasis are most active at night and breed in wet soil, organic matter, or animal burrows [3]. In Colombia, 10 out of the 20 species that can infect both humans and other living beings are present. The cutaneous leishmaniasis (CL) form is the most frequent (98–99%), with the population under five years old and immunocompromised individuals being the most affected [4,5].

The number of CL cases reported in Colombia in 2022 was 4906, with the departments of Amazonas, Boyacá, Caquetá, Cesar, Córdoba, Cundinamarca, Putumayo, Santander, and Sucre being the most affected areas [6].

Since the late 1980s, *Leishmania*-HIV co-infection has been reported in 35 countries, and there have also been other cases of *Leishmania*-Malaria co-infection, which are associated with the worsening of the clinical condition of patients with leishmaniasis. This co-infection type has increased the disease's burden due to the greater difficulty of clinical treatment [7,8]. Currently, antimonial compounds are the primary treatment for leishmaniasis; however, they present high toxicity and resistance in some endemic regions. To address these challenges, alternative drugs have been developed, such as liposomal amphotericin B, which significantly reduces the side effects and treatment duration associated with free amphotericin B but is expensive [9,10]. Other drugs, such as paromomycin and miltefosine, have been associated with high toxicity, resistance, and teratogenic and abortive effects, promoting the discovery and development of low-cost, highly effective drugs with low toxicity [11]. Furthermore, it is worth noting that while *Leishmania* is a parasitic disease mainly affecting humans, it also affects animals such as dogs and rodents, which can serve as reservoirs for the parasite and increase the risk of transmission to humans [12,13].

Therefore, efforts to develop effective treatments and control measures must be considered. High-throughput screening (HTS) has been used since the early 1990s to test the activity of large numbers of molecules against different diseases and thereby identify potential hits for drug development [14]. However, the uncertainty of success, as well as the time and screening costs, limit the use of this technique [15]. In recent years, chemoinformatics tools (e.g., molecular docking, machine learning) have been utilized to conduct *in silico* studies that can predict the interactions between a protein and a ligand, reducing the number of actual laboratory experiments and accelerating the drug discovery process more efficiently and cost-effectively [14,16]. The different research conducted in this field has led to the development of increasingly efficient and better classifying models, which take advantage of large compound databases, opening the possibility of studying diseases that mainly affect poorer populations (NTD), which are not attractive to large industries and big pharma [17].

Leishmaniasis is commonly treated with plants from the Asteraceae family in traditional medicine. Given the diversity of this family (32,913 species) and the wide range of phytochemicals they contain, including alkaloids, coumarins, flavonoids, benzofurans,

sterols, and terpenoids, they are considered a promising source of new leishmanicidal compounds [18]. Some secondary metabolites studied in this family have been sesquiterpenoids [19,20], triterpenes [21], phytosterols [22], and kauranes [23]. However, although they have shown activity to inhibit the disease, their pIC_{50} is not large enough, and compounds that are effective at low concentrations and selective against the parasite are preferred. A group of compounds that has not yet been studied, with records reporting promising *in vitro* activity, is the derivatives of cinnamic acid belonging to the Asteraceae family [24–26].

Gouri et al. report some natural inhibitors against *Leishmania* amastigotes, such as luteolin ($IC_{50} = 3.12 \mu M$), quercetin ($IC_{50} = 10.5 \mu M$), chrysin ($IC_{50} = 13 \mu M$), apigenin, myricetin, cinnamic acid ($IC_{50} = 0.25 \mu M$), and licochalcone A ($IC_{50} = 0.9 \mu M$), which can play an important role in drug discovery [24]. Peixoto et al., on the other hand, evaluated the biological activity of 25 cinnamic acid derivatives against *Leishmania braziliensis* amastigotes, obtaining promising results and finding that aromatic rings with oxygen as a heteroatom had a beneficial effect in terms of activity against *Leishmania* [25]. Considering that heterocyclic compounds have been of great importance for drug development in the pharmaceutical industry, derivatives of cinnamic acid, which is an aromatic carboxylic acid commonly substituted in the *trans* position by an acrylic acid group, represent an interesting starting point for directing studies in the search for possible hits against different species of leishmaniasis [27]. Although some of these compounds have already been studied, many more remain to be analyzed.

Some cinnamic acid derivatives, such as indole-based inhibitors with a Michael acceptor cinnamic ester head, have been tested against human coronaviruses, demonstrating EC_{50} values of $9.14 \mu M$ and $10.1 \mu M$ [28]. Another area in which their potential has been demonstrated is as antitumor agents. In this context, it has been found that brefeldin A 4-*O*-(4)-dimethylaminocinnamate improves aqueous solubility and exhibits strong cytotoxic activity against HepG2 and BEL-7402 cell lines, with IC_{50} values of 0.29 and $0.84 \mu M$, respectively [29].

Additionally, the compound (*E*)-*N*-(2-(dimethylamino)ethyl)-3-(1*H*-indol-3-yl)-*N*-(pyridin-2-yl) acrylamide has shown promise as a focal adhesion kinase (FAK) inhibitor for the intervention in metastatic triple-negative breast cancer. It potently inhibits the proliferation, invasion, and migration of TNBC cells *in vitro*, with an IC_{50} of $8.37 \mu M$ [30]. Additionally, these types of compounds have been proven to be potential anti-inflammatory agents by inhibiting Akt/NF- κ B and MAPK signaling pathways. Among

them, ursodeoxycholic acid–cinnamic acid hybrids showed the best inhibitory activity, with an IC_{50} of 7.70 μ M and no significant toxicity [31].

In the present study, a computational approach was undertaken to identify potential inhibitors of the bifunctional enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS) of *Leishmania major* given its crucial role in the synthesis of DNA in trypanosomatids, which is essential for the parasite's reproduction [32]. To accomplish this, a custom-made, in-house library containing 314 specialized metabolites derived from cinnamic acid was virtually screened.

Initially, a ligand-based predictive classification model was developed using experimental information on the IC_{50} values retrieved from *in vitro* assays of reported compounds against *Leishmania*. Simultaneously, employing a hybrid *Lm*DHFR-TS model constructed based on its amino acid sequence [33], a structure-based ranking through molecular docking calculations was performed using the investigated specialized metabolite database. Through a consensus analysis, molecules with the highest probability of being inhibitors by both approaches were classified as possible hits.

These secondary metabolites were further evaluated through *in vitro* assays using the recombinant *Lm*DHFR-TS, and ADMET properties were calculated to determine their pharmacokinetic properties.

2. Results and Discussion

2.1. Combined Ligand-/Structure-Based Virtual Screening Approach Using *Lm*DHFR-TS.

2.1.1. Ligand-Based Virtual Screening

Initially, a compilation of compounds exhibiting inhibitory activity against *Lm*DHFR-TS was assembled from the ChEMBL database. These compounds underwent classification as either active or inactive, a determination based on their reported IC_{50} values. A cutoff point of $pIC_{50} = 5.0$ was employed for this classification. The choice of this threshold was grounded in the range of IC_{50} values documented in the ChEMBL database, with an effort to strike a balance between the number of active and inactive compounds. This specific value aims to optimize the representation of chemical space for both active and inactive structure classes while concurrently minimizing the false positive rate of the model.

To refine the dataset, duplicate molecules were eliminated during the data curation process, ensuring the creation of a virtual screening model characterized by high prediction efficiency. Additionally, molecules with an IC₅₀ value falling within ± 0.1 of the cut-off point were included in the analysis. Ultimately, a total of 790 molecules were chosen for model training. Within this set, 378 were identified as inactive (47.8%), while 412 were recognized as active (52.2%).

In the ligand-based process, VolSurf+ (128) and AlvaDesc (more than 4000) molecular descriptors were calculated from the three-dimensional representation of each compound in the database. For AlvaDesc molecular descriptors, a feature selection was conducted before model training. This process involved removing all constant variables, variables with only one unique value, variables that had at least one sample with a missing value or exhibited autocorrelation greater than 0.95. After this process, 523 molecular descriptors were used for the model construction.

These descriptors were then utilized to construct the random forest (RF) model in Knime software (KNIME 4.5.0, the Konstanz Information Miner, Copyright 2003–2014, www.knime.org (accessed on 2 February 2023)), which comprised 200 decision trees. The Gini index was employed as the split criterion in the RF model to reduce the number of false positive results. The dataset underwent a five-fold cross-validation strategy, where it was divided into five subsets, each containing an 80% modeling set and a 20% validation set. The modeling set was exclusively used for model construction and further subdivided into multiple training and test sets, maintaining an 80%/20% split ratio. These procedures were conducted following the approach described by Fourches et al. [34].

Molecular descriptors play a crucial role in drug discovery and development, serving as representations of the molecular and chemical properties of the compounds under investigation. In this study, the selected descriptors proved to be instrumental. VolSurf+ generates three-dimensional (3D) molecular descriptors based on the distribution of molecular electrostatic potentials and hydrophobicity, encapsulating molecular surface properties, such as size, shape, and electrostatic potential distribution [35,36]. On the other hand, AlvaDesc provides a diverse array of descriptor types, encompassing constitutional descriptors (detailing the number and type of atoms, bonds, and functional groups in the molecule), topological descriptors (representing molecular shape, size, and complexity), electrostatic descriptors (conveying molecular polarity and charge distribution), and quantum mechanical descriptors (pertaining to the electronic structure and properties of the molecule) [37,38].

The performance of the RF model was assessed to compare the efficacy of the two types of descriptors. This assessment included calculating classification precision, recall, F1-score, and Matthew's correlation coefficient (MCC). Additionally, receiver operating characteristic (ROC) curves were analyzed, and the area under the ROC curve (AUC) was calculated (Figure 1). These evaluation metrics are commonly utilized to gauge the effectiveness of binary classification models. ROC curves and their AUCs are frequently employed to evaluate the performance of models that generate continuous output scores or probabilities. AUC serves as a scalar measure of the model's overall ability to distinguish between positive and negative cases [37,39].

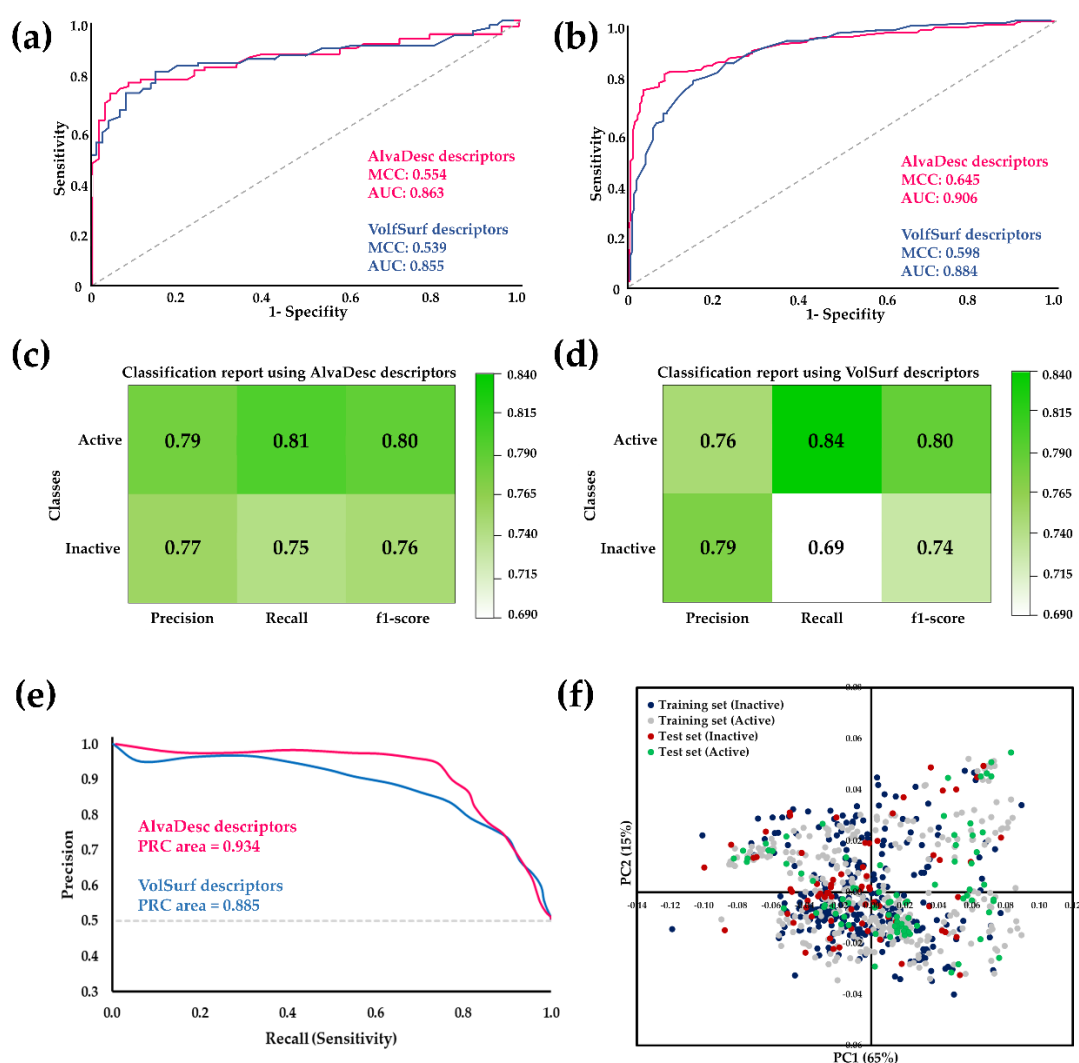


Figure 1. ROC curve comparison for the RF model using AlvaDesc and VolSurf descriptors for (a) test sets and (b) cross-validation. Performance evaluation of RF using (c) AlvaDesc and (d) VolSurf descriptors. (e) Precision–recall (PR) curves for cross-validation. (f) Scatter plots depicting the results of the PCA analysis conducted on the training and test datasets.

According to the parameters presented in Figure 1, it is evident that the MCC and AUC values for both the test sets and cross-validation are higher for AlvaDesc descriptors compared to those obtained for VolSurf descriptors. However, considering that a higher AUC value indicates a more remarkable classification ability of the model and that MCC is expressed in a range of -1 to 1 (where a high value close to 1 suggests a strong correlation between the predicted class and the true class), good values were obtained for both AlvaDesc (AUC: 0.863 and 0.906, MCC: 0.554 and 0.645) and VolSurf (AUC: 0.855 and 0.884, MCC: 0.539 and 0.598) descriptors.

Regarding precision, recall, and F1 score, good and similar values were obtained for both models, except for the recall for inactive compounds in the model created using VolSurf descriptors, which was low, with a value of 0.69. Sensitivity and specificity measures were also calculated to assess the performance of the RF model. For AlvaDesc, the values were 0.807 and 0.752, while for VolSurf, the values were 0.843 and 0.690, respectively. These results indicate a tendency to have few false negatives, a higher value of true negatives, and a lower false positive rate for both descriptors.

The precision–recall (PR) curves, closely related to the ROC curve, were constructed as an evaluation tool for binary classification, enabling the visualization of performance across various thresholds [40]. The results revealed an area under the PR curve of 0.934 for AlvaDesc and 0.885 for VolSurf molecular descriptors, indicating a high-quality model and balanced datasets.

The reliability of the regression model was systematically verified by assessing its applicability domain, ensuring the capability to generate trustworthy predictions. The applicability domain (APD) determination relied on molecular interactions. Results for the training set indicated high reliability rates, reaching 98.1% and 98.4% for the AlvaDesc and VolSurf descriptors, respectively.

Similarly, the test set demonstrated substantial reliability, boasting rates of 96.1% and 100% for the AlvaDesc and VolSurf descriptors, respectively. These results emphasize the model's dependability in predicting outcomes. In the specific context of cinnamic acid derivatives, the APD calculation yielded a noteworthy 80% of structurally reliable outcomes. This analysis further attests to the model's robustness in diverse chemical scenarios.

To enhance insights from the APD and visually represent the chemical space distribution, principal component analysis (PCA) was conducted on the datasets employed in this study. This analysis, performed using the training set, projected the results of the

test set onto the distribution observed for the training set (Figure 1f). Remarkably, the chemical space of the training set encompassed that of the test set, incorporating molecules classified as both active and inactive.

Regarding the model constructed with AlvaDesc molecular descriptors, those demonstrating greater relevance are those associated with the last eigenvector of the Barysz matrix. This can be achieved either by calculating the average of its coefficients (VE2sign_Dz(p)) or by summing them, with the resulting value weighted by the molecule's polarizability (VE2sign_Dz(p)) or by Van der Waals volumes (VE2sign_Dz(v)). Additionally, descriptors AVS_B(m) and AVS_B(v) utilize the charge matrix, summing the elements of a specific row or column, and weighting them by mass or Van der Waals volumes, respectively. Furthermore, descriptors based on extended topochemical atom (ETA) indices are considered, specifically those related to hydrogen bond donor atoms (ETA_D_epsilonD) [37]. The obtained results regarding the relevance of molecular descriptors are presented in Table S1.

The same analysis, evaluating the relevance of molecular descriptors, was also conducted for VolSurf. The two descriptors with the highest accuracy values were associated with the partition coefficient between 1-octanol and water, namely LgD6 and LgD5, which ranked highest. These descriptors calculate the logarithm of the partition coefficient between 1-octanol and water by summing the logP and the fraction of each species at pH 5 and 6, respectively (Table S1).

Additionally, the LogP *n*-oct descriptor emerges as one of the most relevant in model construction, along with LdS5, which computes the logarithm of the partition coefficient between 1-octanol and water through a linear equation derived by fitting GRID-derived atom types to experimental data on *n*-octanol/water partition coefficients. Finally, DD1 appears, measuring the difference between the maximum hydrophobic volumes and the hydrophobic volumes of the imported 3D structure calculated at the first level of energy [35,36].

Ligand-based virtual screening (VS) was utilized to predict the potential inhibitory activity of 314 compounds derived from cinnamic acid in the Asteraceae family, as documented in the literature. Figure 2 showcases the structure and probability of the five best compounds classified using AlvaDesc descriptors. These compounds were (*E*)-2-hydroxy-3',6'-dimethoxychalcone (**103**) [41], apigenin 7-*O*-(6''-caffeoyl)-glucoside (**235**) [42], montamine (**63**) [43], 3-*O*-*p*-coumaroyl-betulinic acid (**150**) [44], and cordoin (**202**) [45]. Additionally, Figure 2 presents the top five compounds predicted using VolSurf

descriptors: 6,8-di-C- β -glucopyranosylchrysin (**242**) [46], montamine (**63**) [43], dihydrocubebin (**305**) [47], prebalanophonin (**312**) [48], and 4-*O*-feruloyl 5-*O*-caffeoylquinic acid (**96**) [49].

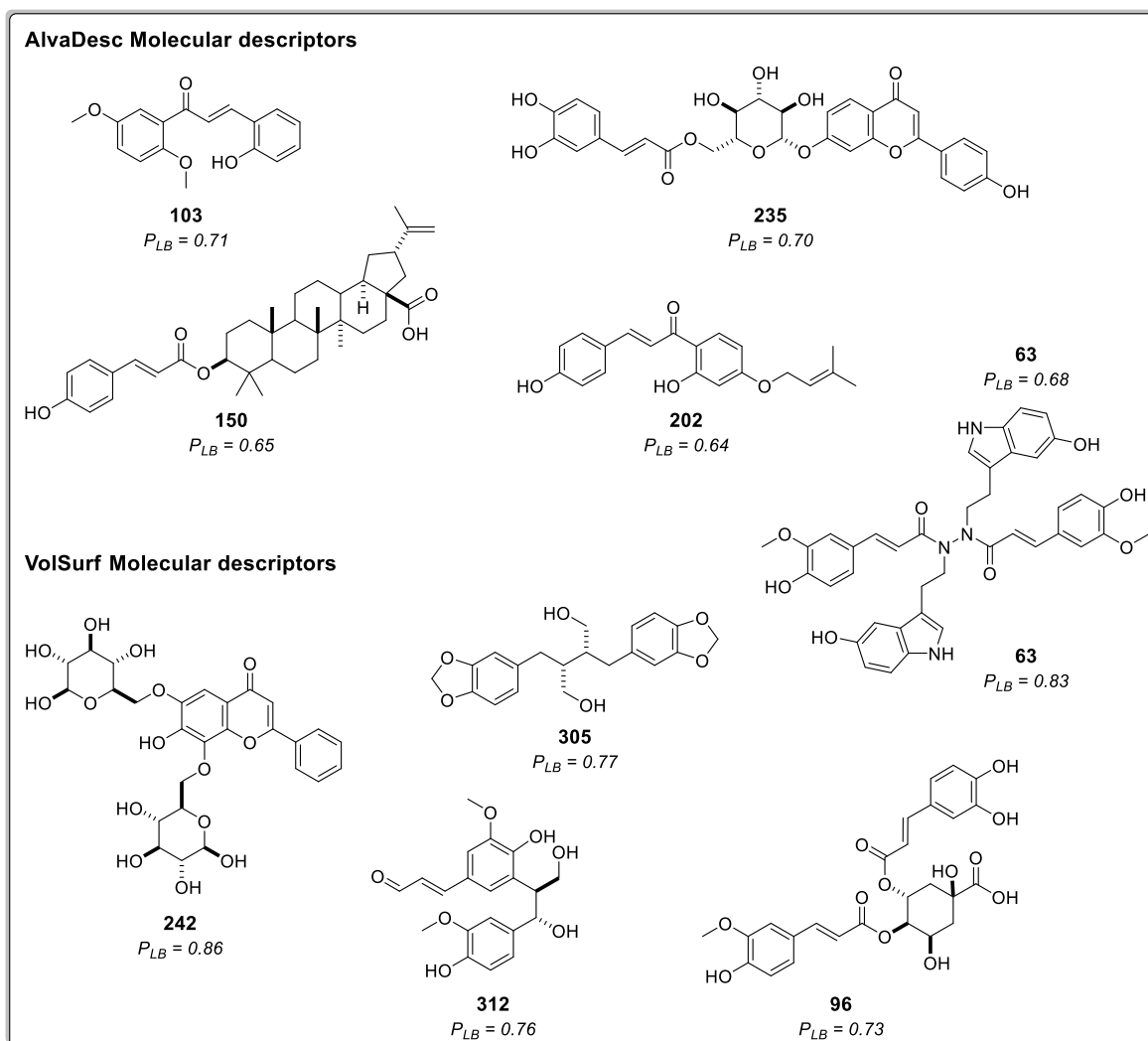


Figure 2. Chemical structures of the five top-ranked cinnamic acid derivatives using a ligand-based virtual screening (LB) with AlvaDesc and VolSurf+ descriptors; P_{LB} = active probability value.

Among all the tested compounds, 116 were classified as active using AlvaDesc molecular descriptors, with probability values ranging from 0.50 to 0.71. On the other hand, 93 compounds were considered active with VolSurf molecular descriptors, and their probability values ranged from 0.50 to 0.86. Some of these molecules were previously reported to exhibit various activities, such as analgesic activity (**305**), antimalarial activity (**150**), cytotoxic activity (**63**), acting as anticancer agents (**202**), and demonstrating antiproliferative properties (**312**) [43,45–47,50–52].

Regarding the best compounds, only one contains nitrogen in its structure (**63**). The rest have various oxygen atoms, forming heterocycles or containing carbonyl groups, ethers, and alcohols. Additionally, one of them is a steroid (**150**), and another is glycosylated (**242**).

2.1.2. Structure-Based Virtual Screening

Structure-based virtual screening (VS) was conducted using a hybrid homology model of *Lm*DHFR-TS [33], a bifunctional enzyme with a critical role in the metabolic pathway of *Leishmania* parasites as well as several protozoa species. The *Leishmania* genus is autotrophic for folate and unconjugated pteridines, with the enzyme DHFR-TS playing a pivotal role in the reduction of dihydrofolate to tetrahydrofolate, a cofactor in the biosynthesis of thymine in nucleotide metabolism [53,54].

The *Lm*DHFR-TS hybrid model was constructed in YASARA software v.19.12.14 and subjected to thorough evaluation for reliability and stereochemical qualities through Ramachandran, WHAT IF, and VERIFY 3D analyses. The Ramachandran plot indicated that 96.9% of residues were in favored regions, confirming model satisfaction (Figure S1). VERIFY 3D results, with 92.6% of residues having a reliable 3D-1D score, and WHAT IF evaluation, showing a mean score of -0.594 , substantiated the model's quality. Dihedral assessment revealed optimal values above 1.085, affirming the robustness of the *Lm*DHFR-TS hybrid model [33].

To assess the potential inhibitory capability of cinnamic acid derivatives against *Lm*DHFR-TS, molecular docking calculations were carried out using Molegro software. The results were validated by redocking the co-crystallized ligand, i.e., ethyl 4-(5-((2,4-diaminoquinazolin-6-yl)methyl)amino)-2-methoxyphenoxy)butanoate (DQ1), along with the reference inhibitor methotrexate (MTX) (Figure 3).

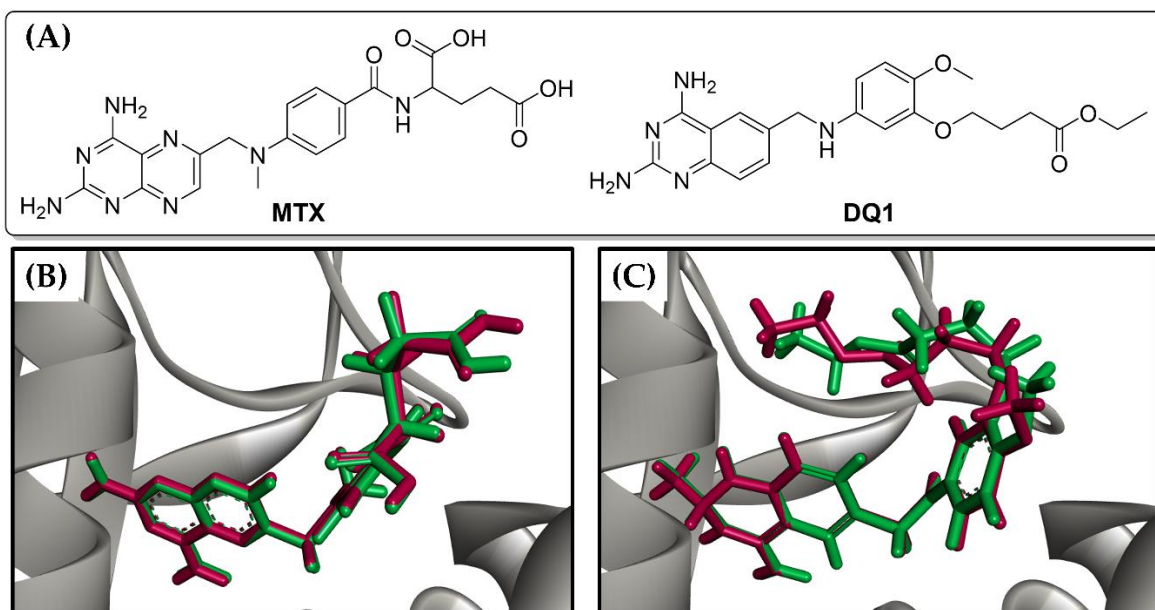


Figure 3. (A) Chemical structures of reference ligands: Methotrexate (MTX) and ethyl 4-(5-[(2,4-diaminoquinazolin-6-yl)methyl]amino)-2-methoxyphenoxy)butanoate (DQ1). Redocking results of (B) MTX and (C) DQ1 in the active site of *LmDHFR-TS*. The original ligand conformation is highlighted in red, while the best pose found in the molecular docking procedure is shown in green.

The compounds were ranked based on the predicted docking binding energy using the probability calculation shown below (Equation (1)), as previously reported by Herrera-Acevedo et al. [19,20]. The ten compounds exhibiting the highest probability of being active are presented in Table 1. Ranked compounds that did not previously show high ligand-based probability values but appeared among the best-ranked derivatives through a structure-based approximation are represented in Figure 4 along with their respective structure-based probability (P_{SB}) values.

$$P_{SB} = (E_i/E_{min}) > 0.5 \text{ and } E_i < E_{ligand} \quad (1)$$

where P_{SB} is the structure-based probability; E_i is the docking energy of compound i , where i ranges from 1 to 314 (cinnamic acid derivatives dataset); E_{min} is the lowest energy value of the dataset; and E_{ligand} is the ligand energy from the co-crystallized inhibitor.

The results showed that the energy-based scoring values were lower for the cinnamic acid derivatives compared to the reference ligands. This suggests that the studied compounds exhibit a higher affinity with the *Lm*DHFR-TS active site in the molecular recognition process. Furthermore, the docking results revealed that 24.5% of the 314 cinnamic acid derivatives dataset had P_{SB} values above 0.5, and among these top-ranked compounds, 64 had a lower docking score than methotrexate, which achieved -114.15 kJ/mol.

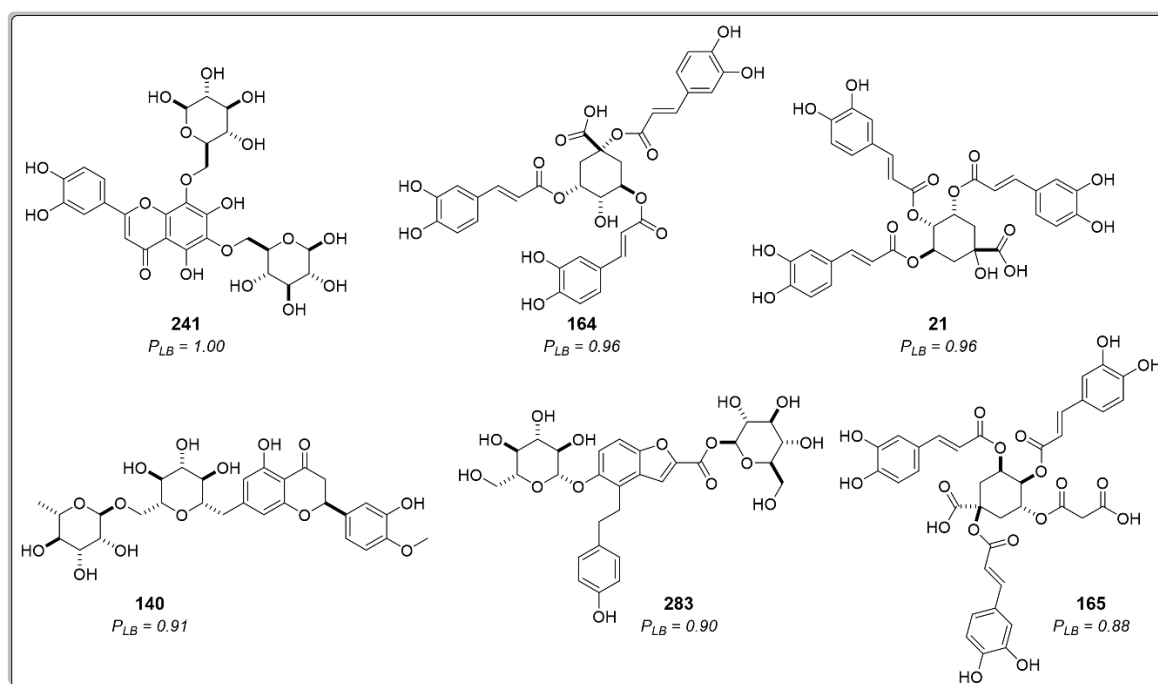


Figure 4. Chemical structure of six of the best-ranked cinnamic acid derivatives that appear as active in the structure-based virtual screening with their respective probability to be active. P_{SB} = active probability value.

Three of the top-ranked molecules predicted to have high ligand-based probability values based on the RF model also demonstrated high structure-based probability values. Specifically, Compound **242**, ranked fourth in the structure-based classification (Table 1), was the best classified in the ligand-based VS model with VolSurf descriptors. Compounds **235** and **63**, positioned among the top ten compounds in structure-based VS with docking scores of -161.4 kJ/mol and -160.1 kJ/mol, respectively, also showed high ligand-based probabilities. Compound **235** was predicted to be the second-best structure with high potential for inhibition using the model built with AlvaDesc descriptors, while Compound **63** was classified in the top three for both RF models (AlvaDesc and VolSurf molecular descriptors).

Table 1. Chemical structure of six of the best-ranked cinnamic acid derivatives that appear as active in the structure-based virtual screening with their respective probability to be active. P_{SB} = active probability value.

Rank	Ligand	Docking Score (kJ/mol)	SD	RMSD
1	241	−182.8	5.4	1.0
2	164	−175.6	7.1	1.8
3	21	−175.5	11.2	1.0
4	242	−169.6	1.9	1.2
5	140	−167.0	3.3	0.4
6	283	−165.4	4.8	1.7
7	165	−161.8	7.4	1.2
8	235	−161.4	5.9	0.9
9	285	−160.9	8.8	1.2
10	63	−160.1	5.2	1.1
Redocking	MTX	−114.2	2.2	0.3
	DQ1	−134.4	2.5	0.3

The analysis of residues for the best poses in the top three compounds revealed that the residues responsible for ligand binding (Val30, Val31, Ala32, Ile45, Trp47, Asp52, Met53, Phe56, Val87, Pro88, Fhe91, Leu94, Val156, Tyr162, and Thr180) have been previously reported in the literature as part of the active site [55]. Certain characteristics of these residues, such as accessibility and charge distribution, enable selective drug design against these protozoans without affecting human enzymes [55]. The interaction diagrams in Figure 5 illustrate that the compound with the highest docking score (Compound **241**, Figure 5C) possesses heterocyclic rings like the reference ligands, with oxygen atoms replacing the nitrogen atoms present in the reference ligands. However, due to the similar electronegativities of nitrogen and oxygen, these atoms favor nearly identical interactions with the enzyme's active site.

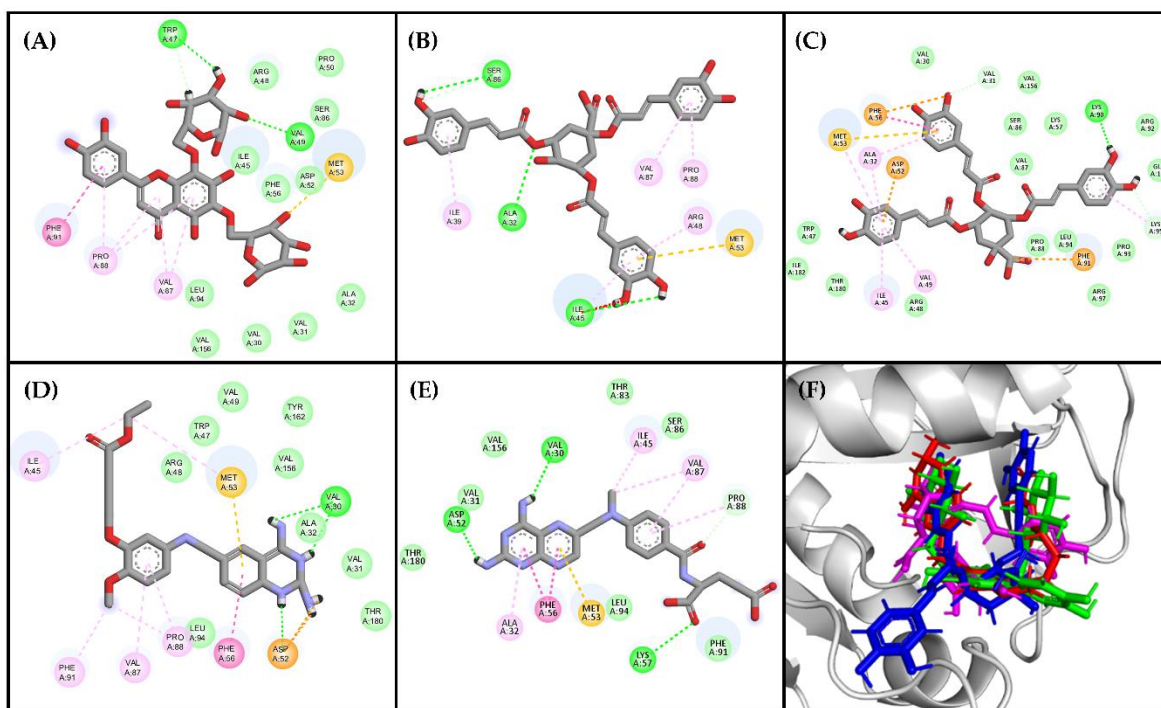


Figure 5. Residual interaction diagrams of (A) compound **241**, (B) compound **164**, (C) compound **21**, (D) DQ1, and (E) methotrexate. Interacting residues are shown in colored circles and dashed lines depending on the type of interaction: H-bond (lime), van der Waals (green), π - π (purple), π -alkyl (pink), unfavorable (red), carbon H-bond (light green), π -anion (orange), π -sulfide (yellowish orange). (F) structural conformations of the coupling between the *LmDHFR-TS* enzyme and the ligands: DQ1 (red), Compound **241** (green), Compound **164** (pink), Compound **21** (blue).

Compounds **164** and **21** lack heterocyclic rings but contain benzene rings, which participate in π - π and π -alkyl interactions. Additionally, these compounds exhibit a relevant number of oxygen-containing groups, such as esters, ethers, and carboxylic acids, facilitating interactions with both residues within the active site and other residues. Specifically, the carboxylic moiety facilitates van der Waals interactions, crucial as they occur with the amino groups in the reference ligands and appear to be important since they are present in the three top-ranked molecules. On the other hand, Compounds **242** and **140**, containing only hydroxyl groups, are less favorable in this binding mode. Although both compounds are isomeric, Compound **164** has few favorable interactions (8 interactions), and Compound **21** has more interactions (25 interactions).

All ligands adopted a U-shaped conformation like the reference ligands DQ1 and MTX (Figure 5F), and most of them formed robust hydrogen bonding interactions with the enzyme (Val156, Val30, Lys95, Met53, Phe91, and Arg97), which are crucial determinants

for binding [53]. To delve deeper into this behavior, a topological polar surface area (TPSA) map was constructed for both the reference ligands and the best-ranked compounds (Figure 6).

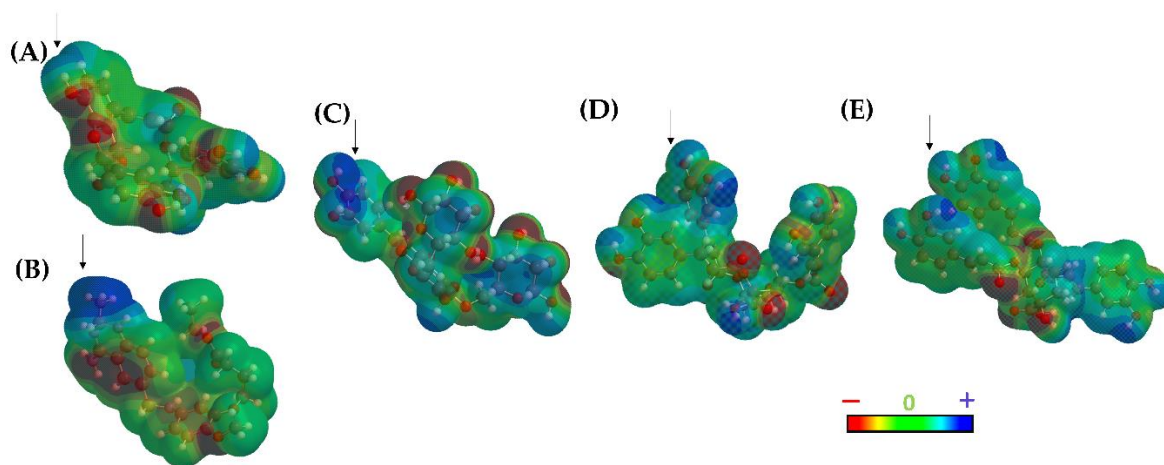


Figure 6. Topological polar surface area (TPSA) map of (A) DQ1, (B) methotrexate, (C) Compound **241**, (D) Compound **164**, and (E) Compound **21**. The arrows show the electron-deficient region of the molecule.

The results of the TPSA maps confirmed a similar spatial distribution among the three top-ranked compounds concerning DQ1 and MTX. An electron-deficient region was identified at the top of the molecule (Figure 6, blue area), which is consistently present in all evaluated molecules, including the two reference ligands. This observation rationalized the similar binding behavior within the active site of *Lm*DHFR-TS, particularly with Met53 as a common crucial contact for these test compounds.

The molecular lipophilic potential (MLP) was also analyzed for both ligands and the protein (2). The results obtained from both TPSA and MLP concerning the active site of *Lm*DHFR-TS show that the active site ends are highly polar, explaining the observed charge distribution in cinnamic acid derivatives.

The lipophilic areas of the pocket predominate in the center of the active site, justifying the charge distribution depicted in Figure 6. Additionally, these calculations revealed a pattern of distribution for polar charges for DQ1, MTX, and the three top-ranked structures. However, this was not observed in the lipophilic regions determined in the MLP. The structure **241** exhibits a pattern like MTX, while Ligands **146** and **21** present lipophilic potential like DQ1 (Figure S2).

2.1.3. Consensus Analysis of the Two VS Approaches

A combined approach was employed to determine the potential activity of cinnamic acids against the *Lm*DHFR-TS enzyme and to mitigate the selection of false positive compounds. This approach incorporated probability scores derived from both structure-based and ligand-based virtual screening (VS) methods in conjunction with the true negative rate obtained from the RF model (Equation (2)) [19].

The design of this approach aimed to assign a higher weight to the ligand-based probability scores (considering their reliance on experimental pIC₅₀ values), in contrast to the structure-based probability scores, which are founded on protein–ligand interactions. This weighting scheme significantly reduces the risk of incorrectly classifying inactive molecules as active (false positives) [23].

$$CA_{Lm} = \frac{P_{SB} + \left((1 + TN_{LB(AD)}) \times P_{LB(AD)} + (1 + TN_{LB(VS)}) \times P_{LB(VS)} \right)}{[3 + TN_{LB(AD)} + TN_{LB(VS)}]} \quad (2)$$

where CA_{Lm} = combined-approach probability, P_{SB} = structure-based probability, TN = true-negative rate, and P_{LB} = ligand-based probability (AD = AlvaDesc descriptors and VS = VolSurf descriptors).

Table 2 presents the results of the best-ranked compounds calculated from the consensus analysis equation. The compounds ranked among the top five for each method are highlighted in bold. Except for **235**, all compounds were classified as potentially active in all virtual screening approximations used in this study. The consensus analysis identified 110 compounds with combined-approach probability values greater than 0.5; however, only 47% of these compounds (**52**) were classified as active through the three *in silico* models used in this study (Table S2). Compound **63** (montamine) was the top-ranked compound. Montamine is an indole alkaloid that has been isolated from Asteraceae species, such as *Centaurea schischkinii* and *Centaurea montana*. Previous studies have reported its anticancer properties [43,56], but its efficacy against *Leishmania* has not been investigated.

Table 2. Cinnamic acid derivatives are classified as active by combining ligand-based and structure-based VS. The numbers in italics represent those compounds classified as active in all three *in silico* models, but they were not previously identified as the best-ranked compounds in any approach.

Rank	Ligand	$P_{LB(AD)}$	$P_{LB(VS)}$	P_{SB}	CA_{Lm}
1	63	0.68	0.83	0.88	0.78
2	242	0.52	0.86	0.93	0.74
3	96	0.55	0.73	0.77	0.67
4	241	0.53	0.55	1.00	0.64
5	39	<i>0.57</i>	<i>0.64</i>	<i>0.77</i>	<i>0.64</i>
6	237	<i>0.61</i>	<i>0.55</i>	<i>0.84</i>	<i>0.64</i>
7	306	<i>0.63</i>	<i>0.53</i>	<i>0.83</i>	<i>0.63</i>
8	165	0.53	0.60	0.88	0.63
9	140	0.59	0.51	0.91	0.63
10	308	<i>0.57</i>	<i>0.59</i>	<i>0.81</i>	<i>0.63</i>

The second best-ranked compound was 6,8-di-*C*- β -glucopyranosylchrysin (**242**), a derivative of chrysin obtained from *Lychnophora ericoides* (Asteraceae). Compared to Compounds **69** (chrysin) and **231** (techtochrysin), classified as inactive, the glycosylated derivative **242** has more hydroxyl groups, enabling interactions with the enzyme's active site. In previous studies, chrysin was biofunctionalized with gold particles due to its low bioavailability, poor absorption, and rapid excretion issues, aiming to neutralize *Leishmania* parasites through its activity against the kinase-3 enzyme [57]. However, Compound **242** could represent an alternative due to its hydrophilic character resulting from the glycosyl groups, potentially inhibiting *Leishmania* parasites by interacting with *LmDHFR-TS*.

The third- and fourth-best-ranked compounds were 4-*O*-feruloyl-5-*O*-caffeoylquinic acid (**96**) and lucenin-2, 6,8-di-*C*- β -glucopyranosylluteolin (**241**), respectively, both extracted from the genus *Lychnophora*—specifically, *Lychnophora ericoides* [46] and *Lychnophora salicifolia* [49], respectively. Additionally, apigenin 7-*O*-rutinoside (**39**), lithospermic acid (**237**), diartigenin (**306**), and isolappaol A (**308**)—four cinnamic acid derivatives that previously exhibited moderate values in both RF models and the molecular docking calculations (all classified as active)—appeared among the top ten

ranked compounds in the combined approach (Figure 7). Hence, these compounds emerge as interesting antileishmanial candidates, as they exhibit activity across all models and maintain consistency in their probability values. Notably, consensus scoring methods are known to enhance hit rates by diminishing the likelihood of false positives [53–58].

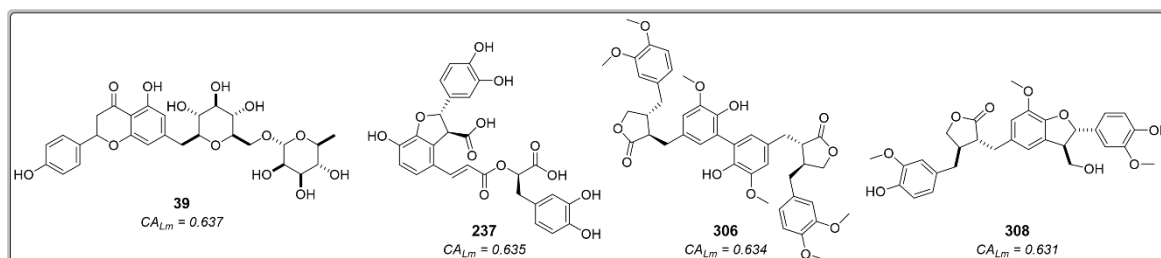


Figure 7. Cinnamic acid derivatives as potential inhibitors of *Lm*DHFR-TS were identified using an approach that combines ligand-based and structure-based virtual screening (VS). CA_{Lm} represents the combined probability value.

The compounds 4-(3,4-dihydroxybenzyl)-2-(3,4-dihydroxyphenyl)tetrahydrofuran-3-carboxy-*O*- β -d-glucopyranoside (**306**) and 7-(3,4-dihydroxyphenyl)-3',4'-dihydroxy-7,8,7',8'-tetrahydronaphtho [8,8'-c]furan-1(3*H*)-one (**308**) are two lignans found in certain species of Asteraceae. Notably, *Hypochaeris radicata* (native to Europe, northern Asia, and parts of North Africa) and *Arctium lappa* (native to Europe and Asia) have been reported as natural sources of these compounds. However, *A. lappa* is widely disseminated in America, and *H. radicata* has also become invasive in regions as far-flung as New Zealand and Chile [59].

Conversely, compound **237**, lithospermic acid, is a common polycyclic phenolic carboxylic acid that has been isolated from species of multiple botanical families, including Lamiaceae and Asteraceae. It has demonstrated a wide range of beneficial properties, acting against cardiovascular diseases and hepatitis. It allows endothelium-dependent vasodilatation, lowers blood pressure, and produces antioxidant effects [60,61].

2.2. Molecular Dynamics Simulations

Conducting molecular dynamics (MD) studies aimed at evaluating protein–ligand stabilities involved considering various factors such as solvent, ions, pressure, and temperature for Compounds **237**, **306**, and **308**. These three compounds emerged as potential inhibitors of *Lm*DHFR-TS based on the consensus analysis of the methodologies employed in this study. Methotrexate (MTX) served as the reference ligand.

The assessment of structural stability was accomplished through root mean square deviation (RMSD) measurements. Over the simulated period of 100 ns, all tested compounds exhibited comparable behavior in relation to the apoenzyme of *LmDHFR-TS* (apo*LmDHFR-TS*, the protein without a ligand) and the *LmDHFR-TS*⋯MTX complex. Upon detailed examination of Figure 8A, it becomes evident that during the initial 30 ns of the simulation, the complexes formed by *LmDHFR-TS* with the three analyzed ligands exhibit behavior like that of the complex with MTX and apo*LmDHFR-TS*. However, after the 40 ns mark, derivatives **237** and **308** display a higher level of disturbance, with RMSD values fluctuating between 0.10 and 0.15 nm (Figure 8A).

Structure **306**, in contrast, maintains behavior like the *LmDHFR-TS*⋯MTX complex throughout the entire 100 ns simulation, with a minor RMSD variation (close to 0.10 nm) compared to the other two analyzed derivatives. This suggests favorable stability of the protein, as the apo*LmDHFR-TS* experiences a variation of 0.15 nm, with a minimum observed at 40 ns and an increase in RMSD values reaching a maximum near 85 ns of the simulation.

Concerning RMSF values (Figure 8B), all examined compounds displayed similar behavior, although specific cases revealed distinct characteristics. Residues Glu218 and Thr410, situated in the protein's loop regions, exhibited the highest fluctuations for the apoenzyme, with Glu218 showing approximately twice the RMSF value compared to the complexes with MTX and the tested cinnamic acid derivatives.

Among the selected compounds, Compound **237** demonstrated higher fluctuations in the loop regions than the other derivatives and MTX, with Gly118, Arg254, and Arg380 being the most variable amino acids. Compounds **306** and **308** exhibited a similar behavior throughout the simulation, showcasing reduced flexibility when complexed.

The critical amino acid residues involved in binding to *LmDHFR-TS* exhibited relatively stable behavior, with RMSF values ranging from 0.10 to 0.20 nm throughout the simulation. Among these residues, Phe91 and Lys95 demonstrated higher variation, exceeding 0.20 nm. In contrast, Arg97 and Val156 exhibited minimal fluctuation, with values close to 0.10 nm. Notably, Val156 in apo*LmDHFR-TS* and the MTX complex displayed lower fluctuation (approximately 30%) compared to the three analyzed cinnamic acid derivatives.

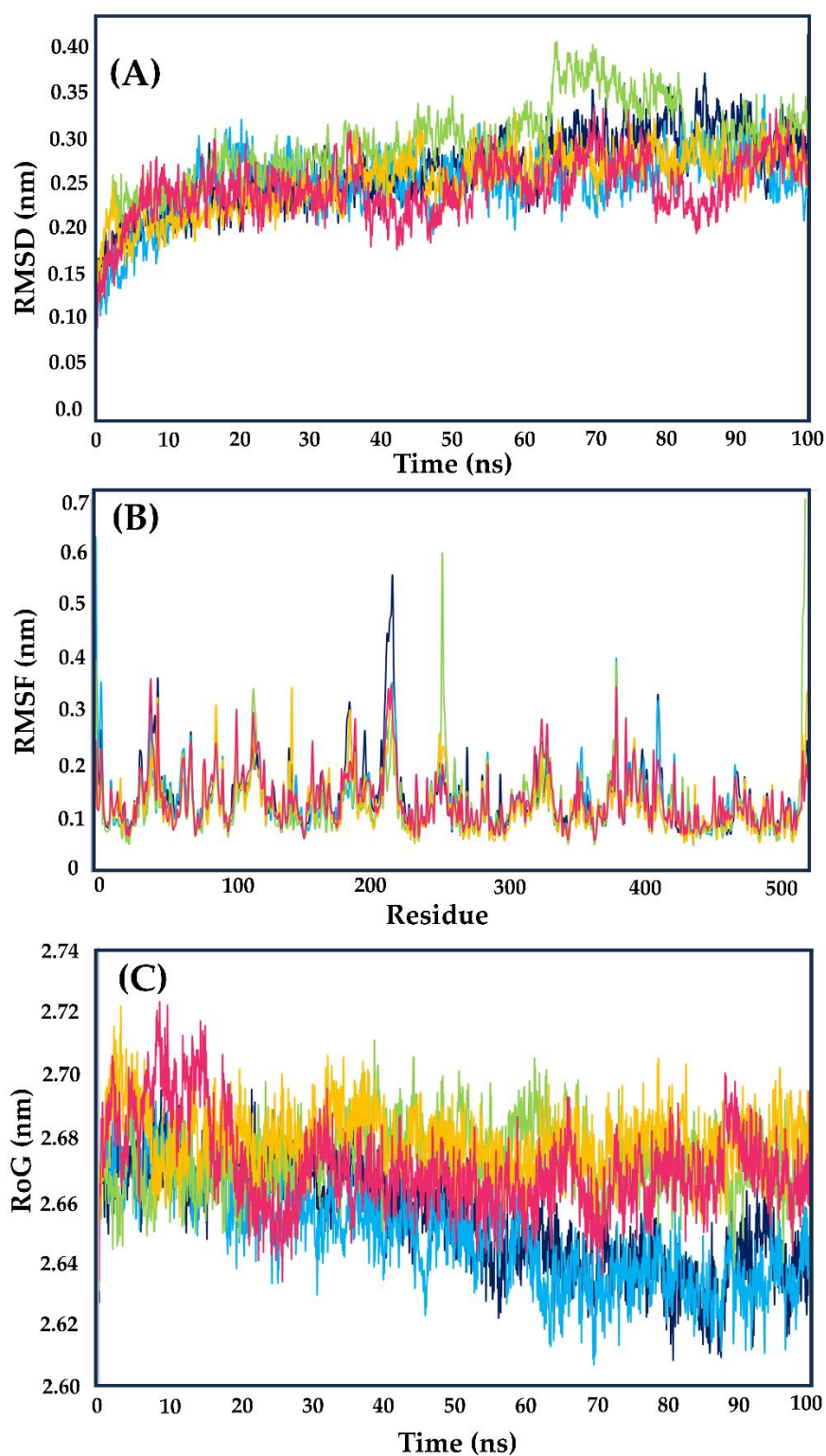


Figure 8. (A) Root mean square deviation (RMSD), (B) root mean square fluctuation (RMSF), and (C) radius of gyration (RoG) values within the LmDHFR-TS binding site obtained after molecular dynamics simulations. Apoenzyme (blue); DHFR-TS...MTX complex (cyan); DHFR-TS...237 complex (light green); DHFR-TS...306 complex (yellow) and DHFR-TS...308 complex (pink).

Conversely, Arg97 displayed values between 0.09 and 0.12 nm. Structure **306** achieved a remarkable value of 0.09 nm, even lower than observed for the MTX complex, while Structures **237** and **208** showed values like those of the apoprotein. Throughout the simulation, the *LmDHFR*-TS complex with Structure **306** consistently promoted protein stability, evidenced by lower RMSF values in this complex, except for residues Leu145 and Lys90.

The structural compactness and mobility of the protein–ligand complexes were assessed throughout the simulation using the radius of gyration (RoG) plot (Figure 8C) [23]. In the initial half of the 50 ns simulation, complexes with cinnamic acid derivatives displayed RoG values indistinguishable from those of the control MTX and apo*LmDHFR*-TS, ranging from 2.64 nm to 2.70 nm.

This indicates a high level of stability and low fluctuations in the tertiary structure. However, after 60 ns, Compounds **237**, **306**, and **308** exhibited similar behavior (varying between 2.64 nm and 2.70 nm) with increased perturbation compared to the DHFR-TS···MTX complex and the apoenzyme, maintaining a consistent mean value with fluctuations ranging from 2.62 to 2.64 nm.

Following molecular dynamic simulations, binding free energies for complexes involving Compounds **237**, **306**, **308**, and the control (MTX) with *LmDHFR*-TS were determined using the MM/PBSA method. The complexes of benzylbutyrolactone-type lignans (**306** and **308**) and the polyphenolic acid (compound **237**) with *LmDHFR*-TS showed binding free energies of −111.1 kJ/mol, −81.0 kJ/mol, and −91.6 kJ/mol, respectively. In all cases, the energy was higher than the −124.5 kJ/mol observed for the complex of MTX with *LmDHFR*-TS (Table 3).

Table 3. Binding free energies (kJ/mol) from the MM/PBSA calculations for Compounds **237**, **306**, and **308** in the active site of *LmDHFR*-TS; MTX was used as the reference ligand.

	237		306		308		MTX	
Energy Contribution	kJ/mol	SD	kJ/mol	SD	kJ/mol	SD	kJ/mol	SD
van der Waals	−218.3	6.2	−209.7	4.6	−217.6	6.2	−239.5	8.2
Electrostatic	−31.3	4.1	−38.0	3.9	−29.0	4.6	−57.3	4.3
Polar solvation	181.5	6.5	157.6	6.3	185.6	6.5	194.6	8.5
SASA	−23.6	1.8	−21.0	1.9	−20.0	1.2	−22.4	2.2
Binding energy	−91.6	4.7	−111.1	4.2	−81	4.6	−124.5	5.8

All complexes under evaluation, including the MTX reference, exhibited a consistent contribution pattern characterized by negative energy values arising from van der Waals, electrostatic, and solvent-accessible surface area (SASA) parameters influencing the binding free energy. The van der Waals parameter, displaying the most substantial negative contribution, registered values lower than −209 kJ/mol. This finding implies that non-polar electrostatic interactions play a pivotal role in the molecular recognition of the *LmDHFR*-TS binding site by the tested compounds.

Concerning polar solvation, all compounds made positive contributions to the total binding energy, with similar values observed for Compounds **237**, **308**, and MTX. Conversely, diartigenin (**306**) exhibited a lesser contribution to this parameter. Additionally, electrostatic interactions negatively influenced the binding free energies, with MTX showing a more significant negative contribution of −57.3 kJ/mol. Meanwhile, the impact of electrostatic interactions for the evaluated cinnamic acid derivatives ranged from 35% to 50% relative to the reference ligand.

2.3. *In Vitro* Enzymatic Activity Inhibition for Selected Cinnamic Acid Derivatives (Compounds **237**, **306**, and **308**) against *LmDHFR*-TS and *HsDHFR*

To validate the outcomes of our combined approach utilizing two virtual screening (VS) methodologies, we conducted *in vitro* enzymatic inhibition assays on five compounds sourced from our in-house library. Compounds **237**, **306**, and **308**, identified as active in all approaches, were selected, along with hesperidin (**140**), a notable flavonoid recognized for its reported antileishmanial activity through apoptosis induction and sterol C-24 reductase

inhibition [62]. Isovitexin 4'-O-glucoside and rutin, demonstrating moderate levels of activity and categorized as inactive in one of the three approaches, were also assessed against *LmDHFR*-TS, with methotrexate serving as the positive control.

The determination of IC₅₀ values involved analyzing concentration-response curves within the 0.1–128 µM range, employing spectrophotometric monitoring of enzymatic activity in a standard DHFR assay. This investigation yielded a spectrum of values ranging from 6.1 to 53.2 µM, corresponding to pIC₅₀ values between 4.27 and 5.21. Notably, Compounds **237**, **306**, and **308** demonstrated the highest activity against *LmDHFR*-TS. Hesperidin (IC₅₀ = 21.6 µM) exhibited substantial activity against the target among the three evaluated flavonoids, with IC₅₀ values of 53.2 µM and 41.7 µM for isovitexin 4'-O-glucoside and rutin, respectively (Table 4).

Table 4. Results of enzymatic activity against *LmDHFR*-TS and *HsDHFR* for selected cinnamic acid derivatives. CI = confidence interval (95%). SI = selectivity index.

Compound	<i>LmDHFR</i> -TS		<i>HsDHFR</i>		SI
	IC ₅₀ (µM)	CI (95%)	IC ₅₀ (µM)	CI (95%)	
hesperidin	21.6	20.2–23.1	86.5	82.3–87.2	4.0
lithospermic acid (237)	7.5	6.8–7.9	22.6	21.3–24.7	3.0
diarctigenin (306)	6.1	5.7–6.4	27.9	26.8–28.6	4.6
isolappaol A (308)	10.1	9.7–10.3	44.8	42.4–45.9	4.4
isovitexin 4'-O-glucoside	53.2	51.1–54.1	125.7	122.8–127.8	2.4
rutin	41.7	40.3–43.1	188.9	186.2–190.6	4.5
MTX	1.4	1.1–1.5	4.9	4.7–5.1	3.5

Structurally, we sought to establish a correlation between the inhibitory activity against *LmDHFR*-TS and the interaction of hydrogen bond acceptors and donors, particularly carbonyl and hydroxyl groups. Among the lignans—**306** and **308**—the presence of the γ-butyrolactone moiety highlighted that the most active compound (**306**) possessed a higher number of carbonyl groups compared to **308**—a feature shared with lithospermic acid (**237**). However, the glycosylated flavonoids (hesperidin, isovitexin 4'-O-glucoside, and rutin) exhibited low inhibitory activities, suggesting that the abundant hydroxyl groups may negatively impact inhibitory activity.

Following this, we calculated the selectivity index (SI) based on the results obtained from *in vitro* tests using the recombinant protein of *Homo sapiens* (*Hs*) DHFR.

The IC₅₀ values against HsDHFR revealed a distinct pattern, implying different mechanisms of action for these two proteins. Moderate SI values were observed, with both benzylbutyrolactone-type lignans (Compounds **306** and **308**) exhibiting the highest SI values—4.6 and 4.4, respectively. Notably, both lignans demonstrated higher SI values than MTX, employed as a positive control (Table 4)

2.4. Pharmacokinetic Properties Predictions

The pharmacokinetic properties, encompassing absorption, distribution, metabolism, excretion, and toxicity (ADMET), of Compounds **237**, **306**, and **308** were predicted using ADMETlab 2.0 and OSIRIS DataWarrior 5.5.0 [63,64]. Multiple approaches were employed to evaluate oral bioavailability, yielding mixed results. While all compounds adhered to Lipinski's "rule of five" [65], none met the criteria set by Pfizer [66] and GSK [67], suggesting potential challenges in oral bioavailability (Table S3).

Regarding cytochrome P450 (CYP) and its isoenzymes, compound **237** exhibited a significant probability of inhibiting CYP2C9. Similarly, Compounds **306** and **308** demonstrated potential inhibition of CYP2C19, CYP2C9, and CYP3A4, indicating potential impacts on the metabolism of other drugs. Conversely, Compound **237** was predicted to act as a substrate for CYP2C9, while Compounds **306** and **308** were associated with CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4, suggesting that they could be metabolized by these isoenzymes. Furthermore, none of the studied compounds exhibited mutagenic, tumorigenic, reproductive, or irritant effects. Identifying potential hERG channel blockers is crucial for assessing the risk of cardiotoxicity [68], and for the three structures, the probabilities of hERG blocking were at most 0.212.

3. Materials and Methods

3.1. Cinnamic Acid Derivatives In-House Dataset

A custom-made, in-house virtual library of 314 distinct cinnamic acid derivatives was built from 76 scientific articles using various search criteria, including keywords such as Asteraceae, Cinnamic Acid Derivatives, Lignans, Polyphenols, Flavonoids, and others. ChemAxon MarvinSketch (ChemAxon, version 21.18.0 (2021), a calculation module developed by ChemAxon, <https://www.chemaxon.com/>, accessed on 12 January 2023) was used to design all the structures.

The three-dimensional (3D) structures for the entire set were generated using Standardizer software (JChem, version 21.18.0 (2021), a calculation module developed by ChemAxon, <https://www.chemaxon.com/>, accessed on 12 January 2023). This software standardized the structures, added hydrogens, performed aromatic form conversions, and refined molecular graphs in three dimensions. The process employs a divide-and-conquer strategy, wherein the structure is partitioned into smaller fragments. These fragments are then organized into a tree based on connectivity information. Conformers generated for the initial structure, represented by the root node in the tree, undergo optimization. The tree-building process incorporates a proprietary extended version of the Dreiding force field [69]. The final dataset was saved in special data file (.sdf) format.

3.2. Classificatory Machine Learning Models

The analyses described below utilized Knime 4.5.0 software (KNIME 4.5.0, the Konstanz Information Miner, Copyright 2003–2014, www.knime.org (accessed on 2 February 2023)) [70]. The process commenced with importing those descriptors generated by the Volsurf+ [35,36] and AlvaDesc [37,38] programs in CSV format.

Subsequently, these descriptors underwent segmentation via the “Partitioning” node, implementing the stratified sampling option, with 80% of the initial dataset designated as the training set and the remaining 20% composing the test set. Random splits were also explored while maintaining consistent ratios for both training and test sets.

The model’s creation processes entailed utilizing the modeling set and the RF algorithm, executed through a five-fold cross-validation procedure employing WEKA nodes. This approach provides a robust and efficient means to evaluate a model’s performance by partitioning the data into five subsets for testing and training, facilitating model selection and generalization assessment [23].

The applicability domain was assessed through Euclidean distances, targeting compounds in the test set with potentially unreliable predictions. A compound was considered unreliable if its applicability domain value exceeded $d + Z\sigma$, where d represents the average Euclidean distance, and σ is the standard deviation of the samples in the training set. These samples exhibited Euclidean distance values lower than the average when compared to all training set samples, with Z serving as an empirical cutoff value set at 0.5 by default [20,71].

To complement these findings and provide a more comprehensive visualization of the chemical space within the datasets used for model construction, principal component

analysis was conducted on the four datasets, encompassing both active and inactive structures for both the training and test sets. This analysis was executed using Unscrambler X (The Unscrambler® X v10.3 User Manual Version 1.0 CAMO SOFTWARE AS, Oslo, Norway).

The RF models were fine-tuned with 200 trees and a random number generator seed of 1, and the Gini index was utilized as the split criterion for both the training and cross-validation sets. These parameter choices were informed by a thorough evaluation of relevant hyperparameters for the machine learning model. The “number of trees” parameter was explored across a range from 100 to 1000, with 200 trees identified as the optimal selection for achieving the best quality parameters. Subsequently, the Gini index was meticulously chosen as the preferred split criterion (Table S4).

Performance analysis of the selected models encompassed an evaluation of both internal and external aspects, incorporating parameters such as sensitivity (true-positive rate), specificity (true-negative rate), and accuracy (overall predictability), derived from the confusion matrix. To offer a more comprehensive understanding of the model’s performance beyond accuracy, the ROC curve was employed. Generated through an “ROC curve” node, this curve relies on sensitivity and specificity. The AUC values derived from the ROC curve range from 0.5, indicating an inability to distinguish between the two groups, to 1, signifying perfect separation without overlap [72]. Additionally, the Matthews correlation coefficient (MCC) was calculated, in which a value of 1 represents perfect prediction, 0 denotes random prediction, and -1 indicates complete disagreement between prediction and observation [73].

Moreover, a performance evaluation of the RF model using AlvaDesc and VolSurf+ descriptors was conducted. This evaluation included precision, recall, and F1 score metrics for both active and inactive sets.

3.3. Molecular Docking Calculations

Molecular docking calculations involved the hybrid model of *Lm*DHFR-TS bound to methotrexate (MTX) [33] and the three-dimensional structures of the cinnamic acid derivatives. We conducted these calculations using Molegro 6.0.1 software.

To ensure consistency, we removed all water molecules from both the enzyme and compound structures, and we prepared them to use the software’s default settings. The MolDock scoring function was utilized, considering internal ES, internal H-bond, and Sp2–Sp2 torsions as criteria for evaluating the ligands.

The molecular docking process was executed through 10 runs utilizing the MolDock SE algorithm. It allowed for a maximum of 1500 interactions, maintained a population size of 50, included up to 300 steps, employed a neighbor distance factor of 1.00, and returned a maximum of 5 poses. To cover the enzyme's ligand-binding site, we established a grid with a 15 Å radius and 0.30 Å resolution [23,33].

Our results were categorized according to docking scores, reflecting the free energy or affinity of the interactions. Each calculation was repeated three times to ensure reliability. For comparison, we employed methotrexate (MTX) as a control.

Topological polar surface area (TPSA) maps were calculated using Spartan 14 for Windows Spartan'14 (Wavefunction Inc., Irvine, CA, USA) [74]. Molecular lipophilic potential (MLP) maps for ligands were calculated in Molinspiration (Molinspiration, Cheminformatics free web services, <https://www.molinspiration.com> (accessed on 24 November 2023), Slovensky Grob, Slovakia). For *LmDHFR*-TS, MLP and TPSA were calculated using ChimeraX [75]. The visualization of two-dimensional residual interaction diagrams was accomplished using Discovery Studio Visualizer v21.1.0.20298 (BIOVIA, Dassault Systèmes, San Diego, CA, USA) [23,33].

3.4. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were conducted in YASARA Structure v. 19.12.14 [76], employing the AMBER14 force field to model the enzyme and ligand–enzyme systems. Before the simulations, each protein underwent hydrogen bond optimization, and chloride (Cl^-) and (Na^+) ions were added to the model systems through the transferable intermolecular potential 3-point (TIP3P) employing 0.997 g/L density for solvating the simulation cell. Acid dissociation constant values (pKa) were calculated for enzymes' titratable amino acids with the H-bonding network and the side-chain placement using a rotamer library (SCWRL) algorithm. Periodic boundary conditions were applied to facilitate the simulations, involving a cell size set 10 Å larger than the protein's size in all instances.

An initial 5000-cycle energy minimization step was carried out using the steepest gradient approach. MD simulations used the particle-mesh Ewald (PME) method to account for long-range electrostatic interactions (8-Å cut-off distance). The simulations were performed under physiological conditions at 298 °K, pH 7.4, and 0.9% NaCl. Temperature control was maintained using a Berendsen thermostat while keeping the

pressure constant. A multiple-time step algorithm with a time step of 2.00 fs was employed.

Finally, MD simulations were run for 100 ns under constant pressure, and the Berendsen thermostat, with snapshots saved at intervals of 100 ps, used the YASARA macro (md_run.mcr) for all simulation phases. Subsequent analyses were also carried out using the default YASARA macro scripts. The molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method was employed to calculate the binding free energies of apoenzyme and enzyme–ligand complexes from the resulting MD trajectories using the g_mmpbsa tool in Gromacs 5.0.5 (open source, <http://www.gromacs.org> (accessed on 17 May 2023)) [77] on an Ubuntu 12.04 server, using NPT and periodic boundary conditions, as previously reported [33,78].

3.5. *LmDHFR*-TS and *HsDHFR* Enzymatic Inhibition Assays

Purification and kinetic characterization of the recombinant *LmDHFR*-TS protein were performed according to the previously reported procedures [33,79,80], while *HsDHFR* protein was obtained from the commercial assay kit (CS0340, Merck KGaA, Darmstadt, Germany). Thus, the *in vitro* evaluation of the top-ranked selected compounds (**237**, **306**, **308**, hesperidin, rutin, and isovitexin 4'-O-glucoside) for inhibitory activity against *LmDHFR*-TS and *HsDHFR* was conducted using a spectrophotometric assay under standard DHFR conditions.

These tested compounds were available from our in-house compound library. Rutin, lithospermic acid, and rutin were commercially purchased (>98%, Merck KGaA, Darmstadt, Germany). Isolappaol and diartigenin were isolated from a commercial *A. lappa* powdered root extract (Prescribed For Life, Fredericksburg, TX, USA) through successive column chromatography, whose spectroscopic data was identical to those of previous reports [81,82].

The assay was conducted with either *LmDHFR*-TS or *HsDHFR* (2.7 nM), bovine serum albumin (BSA, 1 mg/mL), *N*-[tris(hydroxymethyl)-methyl]-2-aminoethanesulfonic acid (TES) buffer (100 mM, pH 7.0, 150 mM β -mercaptoethanol, 2 mM ethylenediaminetetraacetic acid (EDTA)), and nicotinamide adenine dinucleotide phosphate (NADPH, 100 μ M), along with varying concentrations of the test compounds (0.1–128 μ M). The reaction was initiated by adding the substrate (7,8-dihydrofolate (H₂F), 20 μ M) and monitored for 360 s at 340 nm, measuring the oxidation of NADPH to

NADP⁺. This allowed the determination of the initial reaction rate (V_o) through linear regression analysis of the resulting absorbance profile.

All measurements were conducted in triplicate, and methotrexate (MTX) served as a positive control [33]. The resulting V_o values were utilized to calculate the % inhibition, expressed as $100 - (R_i/R_c \times 100)$, where R_i is the V_o in the presence of the inhibitor, and R_c is the V_o in the absence of inhibitors (1% DMSO v/v final concentration). % inhibition was measured for at least five concentrations (0.1–128 μ M) for each test compound (cinnamic acid derivatives and MTX), and concentration-response curves (% inhibition vs. Log[inhibitor]) were constructed using non-linear regression in GraphPad Prism 7.0 (GraphPad, San Diego, CA, USA). [33].

3.6. Pharmacokinetic Properties Predictions

The ADMET parameters for Compounds 237, 306, and 308 were calculated using ADMETlab 2.0, an integrated online platform for predicting ADMET properties [63]. Additionally, drug toxicity predictions were conducted using OSIRIS DataWarrior v.5.2.1, considering parameters such as mutagenicity, tumorigenicity, reproductive effects, and irritability [64].

4. Conclusions

This study identified three cinnamic acid derivatives, lithospermic acid (**237**), diartigenin (**306**), and isolappaol A (**308**), as potential inhibitors of *Lm*DHFR-TS using a combined virtual screening approach (structure/ligand-based). Two random forest models were built using different molecular descriptors. Sensitivity and specificity measures were obtained to evaluate the RF model's performance. The models classified 116 (AlvaDesc) and 93 compounds (VolSurf) as active, showing a tendency to minimize false negatives.

Molecular docking revealed that 24.5% of the 314 cinnamic acid derivatives had values above 0.5, with 64 of them having a lower docking score than methotrexate, the reference ligand. A consensus analysis combining the RF models with molecular docking identified 110 compounds with combined-approach probability values greater than 0.5. From them, 47% were classified as active through the *in silico* models, identifying some compounds with potential leishmanicidal activity that a single approach had not previously highlighted. Lithospermic acid (**237**), diartigenin (**306**), and isolappaol A (**308**) were among the top-ranked compounds, and their binding mode was evaluated using molecular dynamics. Finally, *in vitro* assays using recombinant *Lm*DHFR-TS validated the

computational results, with **237**, **306**, and **308** exhibiting significant activity against *LmDHFR*-TS. However, moderate selective indices (SIs) were observed when assays were performed using *HsDHFR*. Despite this finding, higher SI values than MTX were observed. Thus, these three tested compounds emerged as an interesting alternative as hits against *LmDHFR*-TS; however, specific assays against the parasitic forms of *Leishmania major* are required to extend a clearer prospect for fighting this neglected tropical disease.

References

- [1].Leishmaniasis. Available online: <https://www.who.int/health-topics/leishmaniasis> (accessed on 30 June 2023).
- [2].Abadías-Granado, I.; Diago, A.; Cerro, P.A.; Palma-Ruiz, A.M.; Gilaberte, Y. Cutaneous and Mucocutaneous Leishmaniasis. *Actas Dermosifiliogr. (Engl. Ed.)* **2021**, *112*, 601–618. <https://doi.org/10.1016/j.adengl.2021.05.011>.
- [3].Wijerathna, T.; Gunathilaka, N. Diurnal Adult Resting Sites and Breeding Habitats of Phlebotomine Sand Flies in Cutaneous Leishmaniasis Endemic Areas of Kurunegala District, Sri Lanka. *Parasit. Vectors* **2020**, *13*, 284. <https://doi.org/10.1186/s13071-020-04154-7>.
- [4].Salgado-Almario, J.; Hernández, C.A.; Ovalle, C.E. Geographical Distribution of *Leishmania* Species in Colombia, 1985–2017. *Biomedica* **2019**, *39*, 278–290. <https://doi.org/10.7705/biomedica.v39i3.4312>.
- [5].Ovalle-Bracho, C.; Londoño-Barbosa, D.; Salgado-Almario, J.; González, C. Evaluating the Spatial Distribution of *Leishmania* Parasites in Colombia from Clinical Samples and Human Isolates (1999 to 2016). *PLoS ONE* **2019**, *14*, e0214124. <https://doi.org/10.1371/journal.pone.0214124>.
- [6].Boletín Epidemiológico. Available online: https://www.ins.gov.co/buscador-eventos/BoletinEpidemiologico/2022_Boletín_epidemiologico_semana_25.pdf <https://www.ins.gov.co/buscador-eventos/Paginas/Vista-Boletin-Epidemilologico.aspx> (accessed on 28 May 2023).
- [7].Leishmaniasis—OPS/OMS|Organización Panamericana de la Salud. Available online: <https://www.paho.org/es/temas/leishmaniasis> (accessed on 5 June 2023).
- [8].Wilairatana, P.; Chanmol, W.; Rattaprasert, P.; Masangkay, F.R.; Milanez, G.D.J.; Kotepui, K.U.; Kotepui, M. Prevalence and Characteristics of Malaria Co-Infection among Individuals with Visceral Leishmaniasis in Africa and Asia: A Systematic Review and

- Meta-Analysis. *Parasit. Vectors* **2021**, *14*, 545. <https://doi.org/10.1186/s13071-021-05045-1>.
- [9].Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M.D.F.; Scotti, M.T. Computer-Aided Drug Design Using Sesquiterpene Lactones as Sources of New Structures with Potential Activity against Infectious Neglected Diseases. *Molecules* **2017**, *22*, 79. <https://doi.org/10.3390/molecules22010079>.
- [10].Kumari, S.; Kumar, V.; Tiwari, R.K.; Ravidas, V.; Pandey, K.; Kumar, A. Amphotericin B: A Drug of Choice for Visceral Leishmaniasis. *Acta Trop.* **2022**, *235*, 106661. <https://doi.org/10.1016/j.actatropica.2022.106661>.
- [11].Palić, S.; Beijnen, J.H.; Dorlo, T.P.C. An Update on the Clinical Pharmacology of Miltefosine in the Treatment of Leishmaniasis. *Int. J. Antimicrob. Agents* **2022**, *59*, 106459. <https://doi.org/10.1016/j.ijantimicag.2021.106459>.
- [12].Rafiq, M.; Naveed, M.; Khan, Z.U.; Raza, A.; Awrejcewicz, J.; Soori, A.H.; Ul Haq, I.; Mohsin, M. Modeling the Spread of Leishmaniasis Disease via Delayed Analysis. *Alex. Eng. J.* **2022**, *61*, 11197–11209. <https://doi.org/10.1016/j.aej.2022.05.001>.
- [13].Sánchez-Suárez, J.; Bernal, F.A.; Coy-Barrera, E. Colombian Contributions Fighting Leishmaniasis: A Systematic Review on Antileishmanials Combined with Chemoinformatics Analysis. *Molecules* **2020**, *25*, 5704. <https://doi.org/10.3390/molecules25235704>.
- [14].Hevener, K.E.; Pesavento, R.; Ren, J.; Lee, H.; Ratia, K.; Johnson, M.E. Chapter Twelve—Hit-to-Lead: Hit Validation and Assessment. In *Methods in Enzymology*; Lesburg, C.A., Ed.; Modern Approaches in Drug Discovery; Academic Press: Cambridge, MA, USA, 2018; Volume 610, pp. 265–309. <https://doi.org/10.1016/bs.mie.2018.09.022>.
- [15].Bhattacharjee, A.K. 10—Pharmacophore-Based Virtual Screening of Large Compound Databases Can Aid “Big Data” Problems in Drug Discovery. In *Big Data Analytics in Chemoinformatics and Bioinformatics*; Basak, S.C., Vračko, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2023; pp. 231–246. <https://doi.org/10.1016/B978-0-323-85713-0.00014-1>.
- [16].Shreya; Shweta; Dagur, P.; Rakshit, G.; Ghosh, M. Chapter 7—Virtual Screening of Phytochemicals for Drug Discovery. In *Phytochemistry, Computational Tools and Databases in Drug Discovery*; Egbuna, C., Rudrapal, M., Tijjani, H., Eds.; Drug Discovery Update; Elsevier: Amsterdam, The Netherlands, 2023; pp. 149–179. <https://doi.org/10.1016/B978-0-323-90593-0.00006-X>.

- [17].McGibbon, M.; Money-Kyrle, S.; Blay, V.; Houston, D.R. SCORCH: Improving Structure-Based Virtual Screening with Machine Learning Classifiers, Data Augmentation, and Uncertainty Estimation. *J. Adv. Res.* **2023**, *46*, 135–147. <https://doi.org/10.1016/j.jare.2022.07.001>.
- [18].Joshi, T.; Sharma, P.; Joshi, T.; Mathpal, S.; Pandey, S.C.; Pandey, A.; Chandra, S. Chapter 4—Recent Advances on Computational Approach towards Potential Drug Discovery against Leishmaniasis. In *Pathogenesis, Treatment and Prevention of Leishmaniasis*; Samant, M., Chandra Pandey, S., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 63–84. <https://doi.org/10.1016/B978-0-12-822800-5.00009-3>.
- [19].Acevedo, C.H.; Scotti, L.; Scotti, M.T. In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018**, *13*, 634–645. <https://doi.org/10.1002/cmdc.201700743>.
- [20].Herrera-Acevedo, C.; Dos Santos Maia, M.; Cavalcanti, É.B.V.S.; Coy-Barrera, E.; Scotti, L.; Scotti, M.T. Selection of Antileishmanial Sesquiterpene Lactones from Sistemax Database Using a Combined Ligand-/Structure-Based Virtual Screening Approach. *Mol. Divers.* **2021**, *25*, 2411–2427. <https://doi.org/10.1007/s11030-020-10139-6>.
- [21].Ayeleso, T.B.; Matumba, M.G.; Mukwevho, E. Oleanolic Acid and Its Derivatives: Biological Activities and Therapeutic Potential in Chronic Diseases. *Molecules* **2017**, *22*, 1915. <https://doi.org/10.3390/molecules22111915>.
- [22].Majid Shah, S.; Ullah, F.; Ayaz, M.; Sadiq, A.; Hussain, S.; Ali Shah, A.-U.-H.; Adnan Ali Shah, S.; Wadood, A.; Nadhman, A. β -Sitosterol from *Ifloga Spicata* (Forssk.) Sch. Bip. as Potential Anti-Leishmanial Agent against *Leishmania Tropica*: Docking and Molecular Insights. *Steroids* **2019**, *148*, 56–62. <https://doi.org/10.1016/j.steroids.2019.05.001>.
- [23].Herrera-Acevedo, C.; Flores-Gaspar, A.; Scotti, L.; Mendonça-Junior, F.J.B.; Scotti, M.T.; Coy-Barrera, E. Identification of Kaurane-Type Diterpenes as Inhibitors of *Leishmania* Pteridine Reductase I. *Molecules* **2021**, *26*, 3076. <https://doi.org/10.3390/molecules26113076>.
- [24].Gouri, V.; Upreti, S.; Samant, M. Evaluation of Target-Specific Natural Compounds for Drug Discovery against Leishmaniasis. *Parasitol. Int.* **2022**, *91*, 102622. <https://doi.org/10.1016/j.parint.2022.102622>.
- [25].Rodrigues, M.P.; Tomaz, D.C.; Ângelo de Souza, L.; Onofre, T.S.; Aquiles de Menezes, W.; Almeida-Silva, J.; Suarez-Fontes, A.M.; Rogéria de Almeida, M.; Manoel da

- Silva, A.; Bressan, G.C.; et al. Synthesis of Cinnamic Acid Derivatives and Leishmanicidal Activity against *Leishmania Braziliensis*. *Eur. J. Med. Chem.* **2019**, *183*, 111688. <https://doi.org/10.1016/j.ejmech.2019.111688>.
- [26].Monzote, L.; Perera Córdova, W.H.; García, M.; Piñón, A.; Setzer, W.N. In-Vitro and in-Vivo Activities of Phenolic Compounds against Cutaneous Leishmaniasis. *Rec. Nat. Prod.* **2016**, *10*, 269–276.
- [27].Kabir, E.; Uzzaman, M. A Review on Biological and Medicinal Impact of Heterocyclic Compounds. *Results Chem.* **2022**, *4*, 100606. <https://doi.org/10.1016/j.rechem.2022.100606>.
- [28].Citarella, A.; Moi, D.; Pedrini, M.; Pérez-Peña, H.; Pieraccini, S.; Dimasi, A.; Stagno, C.; Micale, N.; Schirmeister, T.; Sibille, G.; et al. Synthesis of SARS-CoV-2 Mpro Inhibitors Bearing a Cinnamic Ester Warhead with in Vitro Activity against Human Coronaviruses. *Org. Biomol. Chem.* **2023**, *21*, 3811–3824. <https://doi.org/10.1039/D3OB00381G>.
- [29].Jiang, Y.-Y.; Wu, S.; Wu, Y.-W.; Gao, Y.; Chong, D.; Sun, C.; Wei, M.-Y.; Gu, Y.-C.; Shao, C.-L.; Gu, Y. New Brefeldin A-Cinnamic Acid Ester Derivatives as Potential Antitumor Agents: Design, Synthesis and Biological Evaluation. *Eur. J. Med. Chem.* **2022**, *240*, 114598. <https://doi.org/10.1016/j.ejmech.2022.114598>.
- [30].Yang, F.; Xu, K.; Zhang, S.; Zhang, J.; Qiu, Y.; Luo, J.; Tan, G.; Zou, Z.; Wang, W.; Kang, F. Discovery of novel chloropyramine-cinnamic acid hybrids as potential FAK inhibitors for intervention of metastatic triple-negative breast cancer. *Bioorg. Med. Chem.* **2022**, *66*, 116809.
- [31].Li, X.; Hu, Y.; He, B.; Li, L.; Tian, Y.; Xiao, Y.; Shang, H.; Zou, Z. Design, Synthesis and Evaluation of Ursodeoxycholic Acid-Cinnamic Acid Hybrids as Potential Anti-Inflammatory Agents by Inhibiting Akt/NF- κ B and MAPK Signaling Pathways. *Eur. J. Med. Chem.* **2023**, *260*, 115785. <https://doi.org/10.1016/j.ejmech.2023.115785>.
- [32].Sabt, A.; Eldehna, W.M.; Ibrahim, T.M.; Bekhit, A.A.; Batran, R.Z. New Antileishmanial Quinoline Linked Isatin Derivatives Targeting DHFR-TS and PTR1: Design, Synthesis, and Molecular Modeling Studies. *Eur. J. Med. Chem.* **2023**, *246*, 114959. <https://doi.org/10.1016/j.ejmech.2022.114959>.
- [33].Herrera-Acevedo, C.; de Menezes, R.P.B.; de Sousa, N.F.; Scotti, L.; Scotti, M.T.; Coy-Barrera, E. Kaurane-Type Diterpenoids as Potential Inhibitors of Dihydrofolate Reductase-Thymidylate Synthase in New World *Leishmania* Species. *Antibiotics* **2023**, *12*, 663. <https://doi.org/10.3390/antibiotics12040663>.

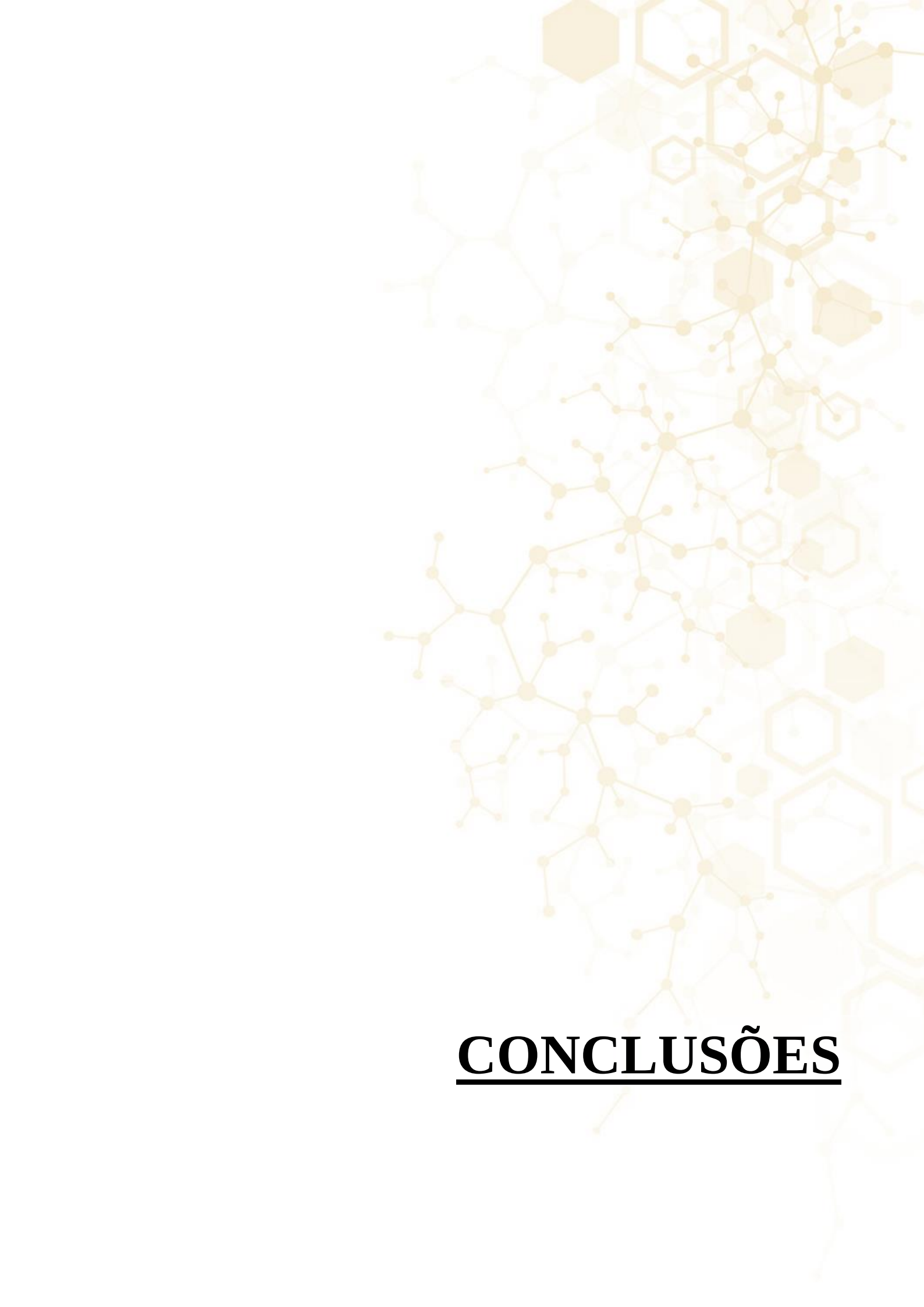
- [34].Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S.Y.; Mumper, R.J.; Tropsha, A. Quantitative nanostructure–activity relationship modeling. *ACS Nano* **2010**, *4*, 5703–5712.
- [35].Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A New Tool for the Pharmacokinetic Optimization of Lead Compounds. *Eur. J. Pharm. Sci.* **2000**, *11* (Suppl. 2), S29–S39. [https://doi.org/10.1016/s0928-0987\(00\)00162-7](https://doi.org/10.1016/s0928-0987(00)00162-7).
- [36].Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular Fields in Quantitative Structure–Permeation Relationships: The VolSurf Approach. *J. Mol. Struct. THEOCHEM* **2000**, *503*, 17–30. [https://doi.org/10.1016/S0166-1280\(99\)00360-7](https://doi.org/10.1016/S0166-1280(99)00360-7).
- [37].Mauri, A. AlvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. In *Ecotoxicological QSARs*; Roy, K., Ed.; Methods in Pharmacology and Toxicology; Springer: New York, NY, USA, 2020; pp. 801–820. https://doi.org/10.1007/978-1-0716-0150-1_32.
- [38].Mauri, A.; Bertola, M. Alvascience: A New Software Suite for the QSAR Workflow Applied to the Blood–Brain Barrier Permeability. *Int. J. Mol. Sci.* **2022**, *23*, 12882. <https://doi.org/10.3390/ijms232112882>.
- [39].Shi, L.; Yan, F.; Liu, H. Screening Model of Candidate Drugs for Breast Cancer Based on Ensemble Learning Algorithm and Molecular Descriptor. *Expert Syst. Appl.* **2023**, *213*, 119185. <https://doi.org/10.1016/j.eswa.2022.119185>.
- [40].Boyd, K.; Eng, K.H.; Page, C.D. Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, 23–27 September 2013*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 13, Part III.
- [41].Ticha, L.A.; Klaasen, J.A.; Green, I.R.; Naidoo, S.; Baker, B.; Pietersen, R.-D. Phytochemical and Antimicrobial Screening of Flavanones and Chalcones from *Galenia Africana* and *Dicerotheramnus Rhinocerotis*. *Nat. Prod. Commun.* **2015**, *10*, 1185–1190.
- [42].Repčák, M.; Krausová, T. Phenolic Glucosides in the Course of Ligulate Flower Development in Diploid and Tetraploid *Matricaria Chamomilla*. *Food Chem.* **2009**, *116*, 19–22. <https://doi.org/10.1016/j.foodchem.2009.01.085>.
- [43].Shoeb, M.; MacManus, S.M.; Jaspars, M.; Trevidu, J.; Nahar, L.; Kong-Thoo-Lin, P.; Sarker, S.D. Montamine, a Unique Dimeric Indole Alkaloid, from the Seeds of *Centaurea montana* (Asteraceae), and Its in Vitro Cytotoxic Activity against the CaCo2 Colon Cancer Cells. *Tetrahedron* **2006**, *62*, 11172–11177. <https://doi.org/10.1016/j.tet.2006.09.020>.

- [44].Zofou, D.; Kuete, V.; Titanji, V.P.K. 17—Antimalarial and Other Antiprotozoal Products from African Medicinal Plants. In *Medicinal Plant Research in Africa*; Kuete, V., Ed.; Elsevier: Oxford, UK, 2013; pp. 661–709. <https://doi.org/10.1016/B978-0-12-405927-6.00017-5>.
- [45].Genovese, S.; Epifano, F.; Curini, M.; Dudra-Jastrzebska, M.; Luszczki, J.J. Prenyloxyphenylpropanoids as a Novel Class of Anticonvulsive Agents. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5419–5422. <https://doi.org/10.1016/j.bmcl.2009.07.110>.
- [46].Gobbo-Neto, L.; Lopes, N.P. Online Identification of Chlorogenic Acids, Sesquiterpene Lactones, and Flavonoids in the Brazilian Arnica *Lychnophora ericoides* Mart. (Asteraceae) Leaves by HPLC-DAD-MS and HPLC-DAD-MS/MS and a Validated HPLC-DAD Method for Their Simultaneous Analysis. *J. Agric. Food Chem.* **2008**, *56*, 1193–1204. <https://doi.org/10.1021/jf072812l>.
- [47].Borsato, M.L.C.; Graef, C.F.F.; Souza, G.E.P.; Lopes, N.P. Analgesic Activity of the Lignans from *Lychnophora Ericoides*. *Phytochemistry* **2000**, *55*, 809–813. [https://doi.org/10.1016/S0031-9422\(00\)00388-5](https://doi.org/10.1016/S0031-9422(00)00388-5).
- [48].Könye, R.; Tóth, G.; Sólyomváry, A.; Mervai, Z.; Zürn, M.; Baghy, K.; Kovalszky, I.; Horváth, P.; Molnár-Perl, I.; Noszál, B.; et al. Chemodiversity of *Cirsium* Fruits: Antiproliferative Lignans, Neolignans and Sesquiterpeneolignans as Chemotaxonomic Markers. *Fitoterapia* **2018**, *127*, 413–419. <https://doi.org/10.1016/j.fitote.2018.04.007>.
- [49].de Athayde, A.E.; Richetti, E.; Wolff, J.; Lusa, M.G.; Biavatti, M.W. “Arnicas” from Brazil: Comparative Analysis among Ten Species. *Rev. Bras. Farmacogn.* **2019**, *29*, 401–424. <https://doi.org/10.1016/j.bjp.2019.02.006>.
- [50].Amoa Onguéné, P.; Ntie-Kang, F.; Lifongo, L.L.; Ndom, J.C.; Sippl, W.; Mbaze, L.M.A. The potential of anti-malarial compounds derived from African medicinal plants. Part I: A pharmacological evaluation of alkaloids and terpenoids. *Malar. J.* **2013**, *12*, 1–26. <https://doi.org/10.1186/1475-2875-12-449>.
- [51].Ahmad, Q.Z.; Jahan, N.; Ahmad, G. Nephroprotective effect of Kabab chini (*Piper cubeba*) in gentamycin-induced nephrotoxicity. *Saudi J. Kidney Dis. Transpl.* **2012**, *23*, 773–781. <https://doi.org/10.4103/1319-2442.98159>.
- [52].Ostad, S.N.; Rajabi, A.; Khademi, R.; Farjadmand, F.; Eftekhari, M.; Hadjiakhoondi, A.; Khanavi, M. Cytotoxic potential of *Centaurea bruguierana* ssp. *belangerana*: The MTT assay. *Acta Med. Iran.* **2016**, *54*, 583–589.
- [53].Maganti, L.; Manoharan, P.; Ghoshal, N. Probing the Structure of *Leishmania* Donovanii Chagasi DHFR-TS: Comparative Protein Modeling and Protein–Ligand

- Interaction Studies. *J. Mol. Model.* **2010**, *16*, 1539–1547. <https://doi.org/10.1007/s00894-010-0649-0>.
- [54].Vickers, T.J.; Beverley, S.M. Folate metabolic pathways in Leishmania. *Essays Biochem.* **2011**, *51*, 63-80. <https://doi.org/10.1042/bse0510063>.
- [55].Zuccotto, F.; Martin, A.C.R.; Laskowski, R.A.; Thornton, J.M.; Gilbert, I.H. Dihydrofolate Reductase: A Potential Drug Target in Trypanosomes and Leishmania. *J. Comput. Aided Mol. Des.* **1998**, *12*, 241–257. <https://doi.org/10.1023/A:1016085005275>.
- [56].Desam, N.R.; Al-Rajab, A.J. Chapter 19—Herbal Biomolecules: Anticancer Agents. In *Herbal Biomolecules in Healthcare Applications*; Mandal, S.C., Nayak, A.K., Dhara, A.K., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 435–474. <https://doi.org/10.1016/B978-0-323-85852-6.00001-9>.
- [57].Raj, S.; Sasidharan, S.; Tripathi, T.; Saudagar, P. Biofunctionalized Chrysin-Conjugated Gold Nanoparticles Neutralize Leishmania Parasites with High Efficacy. *Int. J. Biol. Macromol.* **2022**, *205*, 211–219. <https://doi.org/10.1016/j.ijbiomac.2022.02.047>.
- [58].Onawole, A.T.; Sulaiman, K.O.; Kolapo, T.U.; Akinde, F.O.; Adegoke, R.O. COVID-19: CADD to the rescue. *Virus Res.* **2020**, *285*, 198022.
- [59].Shulha, O.; Çiçek, S.S.; Wangenstein, H.; Kroes, J.; Mäder, M.; Girreser, U.; Sendker, J.; Jöhrer, K.; Greil, R.; Schühly, W.; et al. Lignans and sesquiterpene lactones from *Hypochaeris radicata* subsp. *neapolitana* (Asteraceae, Cichorieae). *Phytochemistry* **2019**, *165*, 112047.
- [60].Begum, T.; Prakash, R.; Duarah, G.; Gogoi, S. Utilization of Rh-carbenoid CH insertion reactions for the synthesis of bioactive natural products. *Stud. Nat. Prod. Chem.* **2020**, *65*, 349–380. <https://doi.org/10.1016/B978-0-12-817905-5.00010-X>.
- [61].Liu, X.; Chen, R.; Shang, Y.; Jiao, B.; Huang, C. Lithospermic acid as a novel xanthine oxidase inhibitor has anti-inflammatory and hypouricemic effects in rats. *Chem. Biol. Interact.* **2008**, *176*, 137–142. <https://doi.org/10.1016/j.cbi.2008.07.003>.
- [62].Tabrez, S.; Rahman, F.; Ali, R.; Akand, S.K.; Alaidarous, M.A.; Banawas, S.; Bin Dukhyil, A.A.; Rub, A. Hesperidin targets Leishmania donovani sterol C-24 reductase to fight against leishmaniasis. *ACS Omega* **2021**, *6*, 8112–8118. <https://doi.org/10.1021/acsomega.0c05858>.
- [63].Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* **2021**, *49*, W5–W14. <https://doi.org/10.1093/nar/gkab255>.

- [64].Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. <https://doi.org/10.1021/ci500588j>.
- [65].Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17. <https://doi.org/10.1016/j.addr.2012.09.019>.
- [66].Hughes, J.D.; Blagg, J.; Price, D.A.; Bailey, S.; DeCrescenzo, G.A.; Devraj, R.V.; et al. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875. <https://doi.org/10.1016/j.bmcl.2008.07.071>.
- [67].Gleeson, M.P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008**, *51*, 817–834. <https://doi.org/10.1021/jm701122q>.
- [68].Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098. <https://doi.org/10.1103/PhysRevA.38.3098>.
- [69].Wang, S.; Li, Y.; Xu, L.; Li, D.; Hou, T. Recent developments in computational prediction of HERG blockage. *Curr. Top. Med. Chem.* **2013**, *13*, 1317–1326. <https://doi.org/10.2174/15680266113139990036>.
- [70].Sun, L.; Yang, H.; Li, J.; Wang, T.; Li, W.; Liu, G.; Tang, Y. In silico prediction of compounds binding to human plasma proteins by QSAR models. *ChemMedChem* **2018**, *13*, 572–581. <https://doi.org/10.1002/cmdc.201700582>.
- [71].Berthold, M.R.; Cebon, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-the Konstanz information miner: Version 2.0 and beyond. *SIGKDD Explor.* **2009**, *11*, 26–31.
- [72].Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36.
- [73].Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451.
- [74].Meng, E.C.; Goddard, T.D.; Pettersen, E.F.; Couch, G.S.; Pearson, Z.J.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **2023**, *32*, e4792. <https://doi.org/10.1002/pro.4792>.
- [75].Hehre, W.J.; Ohlinger, W.A. *Spartan '14*; Wavefunction Inc.: Irvine, CA, USA, 2014; pp. 429–507, ISBN 978-1-890661-42-2.

- [76].Land, H.; Humble, M.S. YASARA: A tool to obtain structural guidance in biocatalytic investigations. In *Protein Engineering; Methods in Molecular Biology*; Springer: New York, NY, USA, 2017; pp. 43–67, ISBN 1064-3745.
- [77].Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- [78].Angarita-Rodríguez, A.; Quiroga, D.; Coy-Barrera, E. Indole-Containing Phytoalexin-Based Bioisosteres as Antifungals: In Vitro and In Silico Evaluation against *Fusarium oxysporum*. *Molecules* **2020**, 25, 45. <https://doi.org/10.3390/molecules25010045>.
- [79].Grumont, R.; Sirawaraporn, W.; Santi, D.V. Heterologous Expression of the Bifunctional Thymidylate Synthase-Dihydrofolate Reductase from *Leishmania major*. *Biochemistry* **1988**, 27, 3776–3784.
- [80].Nare, B.; Hardy, L.W.; Beverley, S.M. The roles of pteridine reductase 1 and dihydrofolate reductase-thymidylate synthase in pteridine metabolism in the protozoan parasite *Leishmania major*. *J. Biol. Chem.* **1997**, 272, 13883–13891.
- [81].Park, S.Y.; Hong, S.S.; Han, X.H.; Hwang, J.S.; Lee, D.; Ro, J.S.; Hwang, B.Y. Lignans from *Arctium lappa* and Their Inhibition of LPS-Induced Nitric Oxide Production. *Chem. Pharm. Bull.* **2007**, 55, 150–152. <https://doi.org/10.1248/cpb.55.150>.
- [82].Umehara, K.; Sugawa, A.; Kuroyanagi, M.; Ueno, A.; Taki, T. Studies on Differentiation-Inducers from *Arctium Fructus*. *Chem. Pharm. Bull.* **1993**, 41, 1774–1779. <https://doi.org/10.1248/cpb.41.1774>.



CONCLUSÕES

Conclusão

O primeiro capítulo deste trabalho fornece perspectivas sobre o design de medicamentos baseado na estrutura do alvo, enfatizando sua importância e as estruturas bem-sucedidas descobertas. Reconhece as limitações e destaca o potencial dos bancos de dados constantemente atualizados e os softwares de docking molecular. O desenvolvimento de funções de pontuação, especialmente para docking flexível de receptores, e avanços em simulações de dinâmica molecular são cruciais para a precisão. Adicionalmente menciona os desafios para a melhoria da precisão dessas metodologias para um design eficaz de medicamentos baseado em alvos.

Na revisão do capítulo II, foi constatado que a utilização de ferramentas de quimioinformática para a pesquisa sobre *Leishmania*, com foco particular nos metabólitos secundários de Asteraceae, tem mostrado promessa na identificação de potenciais candidatos a medicamentos. Embora essas abordagens *in silico* ofereçam perspectivas valiosas, a ausência de resultados *in vitro* correspondentes representa um branco crítico. Estudos futuros devem priorizar validações experimentais para avançar no desenvolvimento de quimioterapias eficazes e seguras contra parasitas da *Leishmania*.

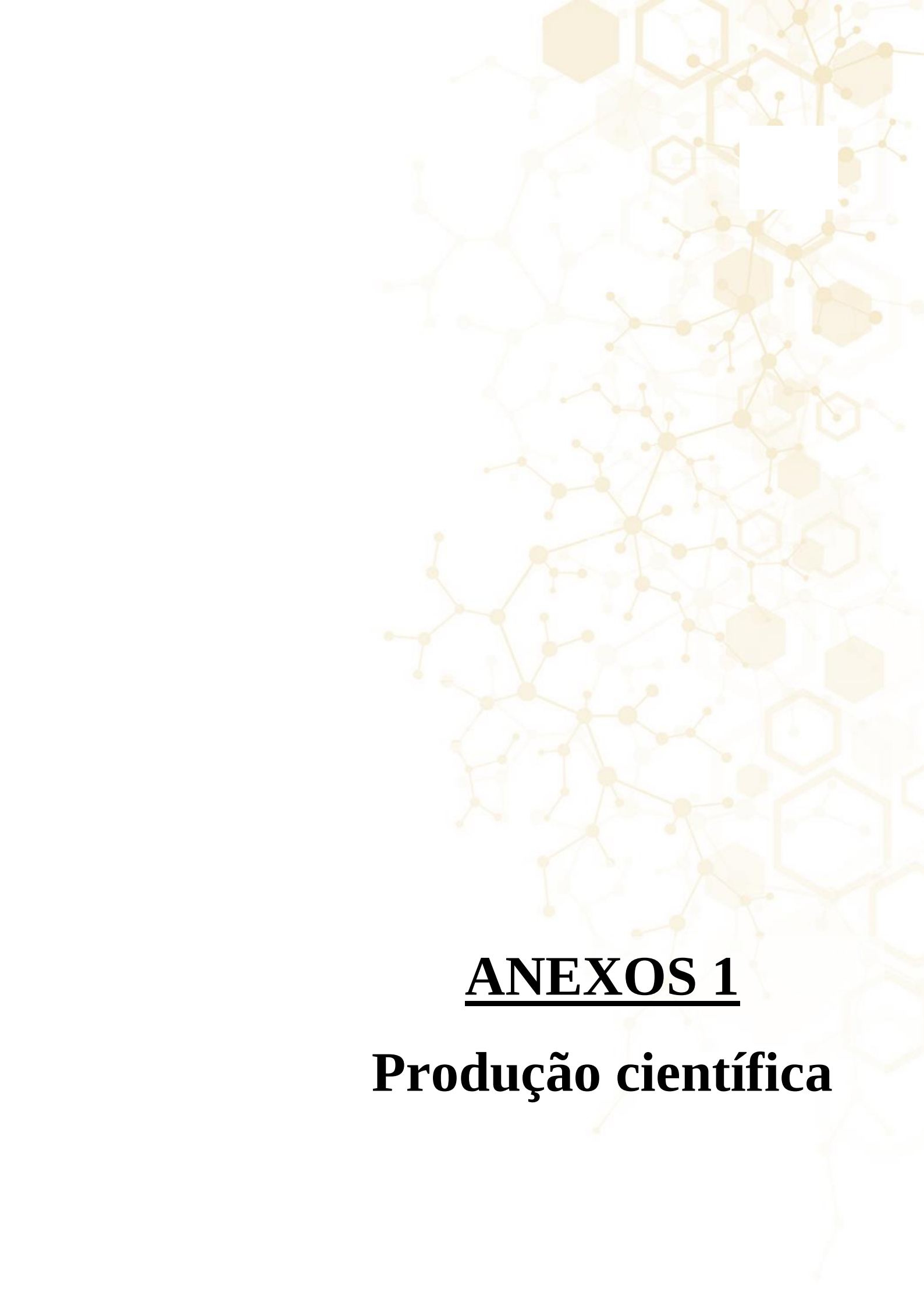
O capítulo III empregou duas abordagens distintas de triagem virtual (VS) para identificar compostos potenciais leishmanicidas partindo de um conjunto de 1306 lactonas sesquiterpênicas (SLs) obtidas do Sistemax. A VS baseada em ligantes, usando dois modelos de random forest (RF), alcançou precisões acima de 71%, revelando características estruturais associadas a amastigotas e promastigotas. A VS baseada em estrutura, usando estruturas cristalinas e um modelo de homologia, identificou SLs com pontuações de docking favoráveis. Uma análise de consenso normalizou as pontuações de probabilidade, revelando 13 potenciais SLs leishmanicidas. Notavelmente, a disecoeudesmanolídeo 3A e a 9 α -linoloyloxi-15-hidroxi-8 β -(2-metilbutiroyloxi)-14-oxo-acantospermolídeo surgiram como potenciais compostos leishmanicidas multialvo. Simulações de dinâmica molecular apoiaram as afinidades com LdPTR1 encontradas. Essa abordagem combinada de VS oferece uma metodologia inovadora para identificar moléculas promissoras e entender seus mecanismos de ação contra *L. donovani*.

Na primeira parte do capítulo IV, os cauranos 135 e 302 surgiram como promissores agentes leishmanicidas contra *L. major*, validados por ensaios *in vitro*. Esses compostos foram identificados a partir de um banco de dados de 360 cauranos usando uma abordagem de aprendizado de máquina e docking molecular. O estudo demonstrou a precisão da abordagem de triagem virtual (VS) e explorou os modos de ligação no local

ativo de *LmPTR1*. Os compostos também exibiram potencial atividade multiespécies contra outras espécies de *Leishmania*, destacando sua importância como compostos líderes para quimioterapias alternativas.

A segunda parte do capítulo identificou os compostos cauranos 3-*p*-cumaroiloxi-ent-caur-16-eno-19-oico e ácido 3-cinamoiloxi-ent-caur-16-eno-19-oico como potenciais inibidores duplos de *PTR1* e *DHFR-TS* em *L. major*, expandindo as descobertas anteriores de inibição de *PTR1*. Ambos compostos exibiram atividade inibitória *in vitro* contra *L. major DHFR-TS*, com valores de IC_{50} de 6,3 e 4,5 μ M, respectivamente. O docking molecular e as simulações de dinâmica confirmaram seu potencial como inibidores multiespécies, oferecendo uma abordagem valiosa para a quimioterapia alternativa contra doenças da *Leishmania*.

Finalmente, no capítulo V, foram identificados o ácido litospérmico, diartigenina e isolappaol A como potenciais inibidores de *LmDHFR-TS* por meio de uma abordagem combinada de triagem virtual. Os cálculos de docking molecular e a análise de consenso revelaram candidatos promissores, com ensaios *in vitro* validando atividade significativa contra *LmDHFR-TS*. Apesar dos índices seletivos moderados nos ensaios com *HsDHFR*, esses compostos mostraram valores de SI superiores ao metotrexato, sugerindo seu potencial como hits contra *LmDHFR-TS*. Pesquisas adicionais com ensaios específicos contra as formas parasitárias de *Leishmania major* são essenciais

The background of the page is a light beige color with a complex, abstract pattern of thin, golden-brown lines and dots, resembling a molecular or network structure. Scattered throughout are several hexagonal shapes, some solid and some outlined, in the same golden-brown color. In the upper right quadrant, there is a white rectangular area that appears to be a placeholder for a logo or image.

ANEXOS 1

Produção científica

1. Natural Products as Potential Agents against SARS-CoV and SARS-CoV-2

Journal: Current Medicinal Chemistry (2021)

Impact Factor: 4.740

Send Orders for Reprints to reprints@benthamscience.net

Current Medicinal Chemistry, 2021, 28, 1-0

1

REVIEW ARTICLE

Natural Products as Potential Agents Against SARS-CoV and SARS-CoV-2

Joanda Paolla Raimundo e Silva¹, Chonny Alexander Herrera Acevedo¹, Thalisson Amorim de Souza¹, Renata Priscila Barros de Menezes¹, Zoe L. Sessions², Lucas Silva Abreu¹, Samuel Paulo Cibulski¹, Luciana Scotti¹, Marcelo Sobral da Silva¹, Eugene N. Muratov², Marcus Tullius Scotti¹ and Josean Fecine Tavares^{1,*}

¹Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa, Brazil; ²Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.

Abstract: Background: Natural products are useful agents for the discovery of new lead-compounds and effective drugs to combat coronaviruses (CoV).

Objective: The present work provides an overview of natural substances, plant extracts, and essential oils as potential anti-SARS-CoV agents. In addition, this work evaluates their drug-like properties which are essential in the selection of compounds in order to accelerate the drug development process.

ARTICLE HISTORY

Received: August 18, 2020
Revised: December 06, 2020
Accepted: December 15, 2020

DOI:
10.2174/0929867328066210125113938

Methods: The search was carried out using PubMed, ScienceDirect and SciFinder. Articles addressing plant-based natural products as potential SARS-CoV or SARS-CoV-2 agents within the last seventeen years were analyzed and selected. The descriptors for Chemometrics analysis were obtained in alvaDesc and the principal component analysis (PCA) was carried out in SIMCA version 13.0.

Results: Based on *in vitro* assays and computational analyses, this review covers twenty-nine medicinal plant species and more than 300 isolated substances as potential anti-coronavirus agents. Among them, flavonoids and terpenes are the most promising compound classes. *In silico* analyses of drug-like properties corroborate these findings and indicate promising candidates for *in vitro* and *in vivo* studies to validate their activity.

Conclusion: This paper highlights the role of ethnopharmacology in drug discovery and suggests the use of integrative (*in silico/ in vitro*) and chemocentric approaches to strengthen current studies and guide future research in the field of antiviral agents.

Keywords: Natural products, drug discovery, SARS-CoV, SARS-CoV-2, COVID-19, *in silico*, *in vitro*.

1. INTRODUCTION

The ongoing severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic, in combination with the previous severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) outbreaks, has shed light on how devastating and life-threatening these emerging viral infections can be [1]. The SARS-CoV-2 outbreak (colloquia

lly COVID-19) was first reported in Wuhan, China, in December 2019. It spreads globally at unprecedented rates, and the World Health Organization (WHO) declared it an international public health crisis on April 05, 2020.

The emergence of SARS-CoV-2 was markedly similar to the 2003 SARS-CoV outbreak that affected 26 countries and resulted in more than 8,000 cases [2]. Both viruses emerged during the winter, transpired from exposure to live animals, and etiological origins in phylogenetically related betacoronaviruses [3]. Similarly, MERS-CoV was a worldwide health concern in 2012 [2]; its high mortality rate affected more than

* Address correspondence to this author at the Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, Postal; Code: 58051-900, João Pessoa, Brazil; E-mail: josean@ltf.ufpb.br

2. Machine learning models to select potential inhibitors of acetylcholinesterase activity from Sistemax: a natural products database.

Journal: Molecular Diversity (2021)

Impact Factor: 3.364

Molecular Diversity
<https://doi.org/10.1007/s11030-021-10245-z>

ORIGINAL ARTICLE



Machine learning models to select potential inhibitors of acetylcholinesterase activity from Sistemax: a natural products database

Chonny Herrera-Acevedo^{1,4} · Camilo Perdomo-Madrigal² · Kenyi Herrera-Acevedo³ · Ericsson Coy-Barrera⁴ · Luciana Scotti¹ · Marcus Tullius Scotti¹

Received: 31 March 2021 / Accepted: 3 June 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Alzheimer's disease is the most common form of dementia, representing 60–70% of dementia cases. The enzyme acetylcholinesterase (AChE) cleaves the ester bonds in acetylcholine and plays an important role in the termination of acetylcholine activity at cholinergic synapses in various regions of the nervous system. The inhibition of acetylcholinesterase is frequently used to treat Alzheimer's disease. In this study, a merged BindingDB and ChEMBL dataset containing molecules with reported half-maximal inhibitory concentration (IC_{50}) values for AChE (7032 molecules) was used to build machine learning classification models for selecting potential AChE inhibitors from the Sistemax dataset (8593 secondary metabolites). A total of seven fivefold models with accuracy above 80% after cross-validation were obtained using three types of molecular descriptors (VolSurf, DRAGON 5.0, and bit-based fingerprints). A total of 521 secondary metabolites (6.1%) were classified as active in this stage. Subsequently, virtual screening was performed, and 25 secondary metabolites were identified as potential inhibitors of AChE. Separately, the crystal structure of AChE in complex with (–)-galantamine was used to perform molecular docking calculations with the entire Sistemax dataset. Consensus analysis of both methodologies was performed. Only eight structures achieved combined probability values above 0.5. Finally, two sesquiterpene lactones, structures **15** and **24**, were predicted to be able to cross the blood–brain barrier, which was confirmed in the VolSurf+ quantitative model, revealing these two structures as the most promising secondary metabolites for AChE inhibition among the 8593 molecules tested.

✉ Marcus Tullius Scotti
mtscotti@gmail.com

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa, PB 58051-900, Brazil

² School of Science, Universidad de Ciencias Aplicadas y Ambientales, Calle 222 # 55–37, Bogotá D.C., Colombia

³ Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Carrera 45 # 26–85, Bogotá D.C., Colombia

⁴ Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, 250247 Cajicá, Colombia

Published online: 16 June 2021

Springer

3. The SistematX Web Portal of Natural Products: An Update.

Journal: Journal of Chemical Information and Modeling

Impact Factor: 6.162



pubs.acs.org/jcim

Application Note

The SistematX Web Portal of Natural Products: An Update

Renan P. O. Costa, Lucas F. Lucena, Lorena Mara A. Silva, Guilherme Julião Zocolo, Chonny Herrera-Acevedo, Luciana Scotti, Fernando Batista Da-Costa, Nikita Ionov, Vladimir Poroikov, Eugene N. Muratov, and Marcus T. Scotti*

Cite This: <https://doi.org/10.1021/acs.jcim.1c00083>

Read Online

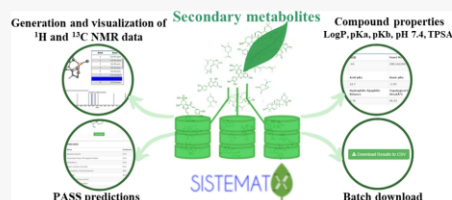
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Natural products and their secondary metabolites are promising starting points for the development of drug prototypes and new drugs, as many current treatments for numerous diseases are directly or indirectly related to such compounds. State-of-the-art, curated, integrated, and frequently updated databases of secondary metabolites are thus highly relevant to drug discovery. The SistematX Web Portal, introduced in 2018, is undergoing development to address this need and documents crucial information about plant secondary metabolites, including the exact location of the species from which the compounds were isolated. SistematX also allows registered users to log in to the data management area and gain access to administrative pages. This study reports recent updates and modifications to the SistematX Web Portal, including a batch download option, the generation and visualization of ^1H and ^{13}C nuclear magnetic resonance spectra, and the calculation of physicochemical (drug-like and lead-like) properties and biological activity profiles. The SistematX Web Portal is freely available at <http://sistematrix.ufpb.br>.



1. INTRODUCTION

Computational approaches have played an increasingly prominent role in natural product (NP)-based drug discovery. For example, the development and use of NP databases allow access to numerous chemical, biological, pharmacological, toxicological, and structural NP data.^{1,2} Sorokina et al. recently reviewed currently available NP databases, citing greater than 120 examples, and noted that access to information was limited, and only a few of the databases were sustainably managed and continually developed.^{2,3} Nevertheless, many NP databases are constantly growing and being updated and are regularly employed in drug design, e.g., ZINC,⁴ which is composed of over 80 000 entries, and SuperNatural II,⁵ the largest NP database currently available.^{3,6}

Free, open NP databases (e.g., the ColleCtion of Open Natural prodUcTs (COCONUT), comprising 406 076 unique, "flat" (lacking stereochemistry) NPs and 730 441 stereochemically preserved NPs) have emerged as important online research tools, facilitating access to NP databases and addressing the limitations observed in many of these computational tools.³ Such NP databases allow for bulk downloads, enabling their use for in silico research, such as virtual screening. Other features, such as the geographical origin of secondary metabolites, have also become increasingly common in NP databases.¹

SistematX (<http://sistematrix.ufpb.br>) is a web portal of natural products developed at the Federal University of Paraíba, PB, Brazil and originally introduced in 2018. It is an open-access database of secondary metabolites available to any

research group. The database initially comprised approximately 2150 secondary metabolites noted across 4000 botanical occurrences of the Asteraceae family, approximately 500 botanical occurrences of the Apocynaceae family, and several terpenes and alkaloids of the Annonaceae family corresponding to greater than 800 botanical occurrences.⁷ SistematX has emerged as a promising tool that connects a large number of molecular properties to the reporting literature, facilitates the use and visualization of these properties, and provides information for chemosystematic studies, compound dereplication, and taxonomic correlations. These goals are achieved through the following features: (a) chemical retrieval by structure, simplified molecular-input line-entry system (SMILES) code, compound name, and plant species; (b) inclusion of chemical structures and characteristics important for NP chemistry in search results; and (c) storage of search results according to the best practices in the field, including curated chemical structures, taxonomy of the plant from which the compound was isolated, bibliographic reference(s), and Global Positioning System coordinates.⁷

Received: January 29, 2021



© XXXX American Chemical Society

A

<https://doi.org/10.1021/acs.jcim.1c00083>
J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

4. Development of 2D, 3D-QSAR and Pharmacophore Modeling of Chalcones for the Inhibition of Monoamine Oxidase B.

Journal: Combinatorial Chemistry and High Throughput Screening (2022)

Impact Factor: 1.800

Send Orders for Reprints to reprints@benthamscience.net

1731

Combinatorial Chemistry & High Throughput Screening, 2022, 25, 1731-1744

RESEARCH ARTICLE



Development of 2D, 3D-QSAR and Pharmacophore Modeling of Chalcones for the Inhibition of Monoamine Oxidase B



Bijo Mathew^{1,*,#}, Chonny Herrera-Acevedo^{2,#}, Sanal Dev³, T.M. Rangarajan⁴, Mohamed Saheer Kuruniyan⁵, Punnoth Poonkuzhi Naseef⁶, Luciana Scotti² and Marcus Tullius Scotti^{2,*}

¹Department of Pharmaceutical Chemistry, Amrita School of Pharmacy, Amrita Vishwa Vidyapeetham, AIMS Health Sciences Campus, Kochi-682 041, India; ²Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900, João Pessoa, Brazil; ³Department of Pharmaceutical Chemistry, Al-Shifa College of Pharmacy, Perinthalmanna - 679322, Kerala, India; ⁴Department of Chemistry, Sri Venkateswara College, University of Delhi, New Delhi-110021, India; ⁵Department of Dental Technology, College of Applied Medical Sciences, King Khalid University, Abha-61421, Saudi Arabia; ⁶Department of Pharmaceutics, Moulana College of Pharmacy, Kerala-679321, India

Abstract: Background: Selective and reversible types of MAO-B inhibitors have emerged as promising candidates for the management of neurodegenerative diseases. Several functionalized chalcone derivatives were shown to have potential reversible MAO-B inhibitory activity, which have recently been reported from our laboratory.

Methods: With the experimental results of about 70 chalcone derivatives, we further developed a pharmacophore modelling, and 2D and 3D-QSAR analyses of these reported chalcones for MAO-B inhibition.

Results: The 2D-QSAR model presented four variables (MATS7v, GATS1i and 3i, and C-006) from 143 Dragon 7 molecular descriptors, with a r^2 value of 0.76 and a Q^2_{cv} for cross-validation equal to 0.72. An external validation also was performed using 11 chalcones, obtaining a Q^2_{ext} value of 0.74. The second 3D-QSAR model using MLR (multiple linear regression) was built starting from 128 Volsurf+ molecular descriptors, being identified as 4 variables (Molecular descriptors): D3, CW1 and LgS11, and L2LGS. Adetermination coefficient (r^2) value of 0.76 and a Q^2_{cv} for cross-validation equal to 0.72 were obtained for this model. An external validation also was performed using 11 chalcones and a Q^2_{ext} value of 0.74 was found.

Conclusion: This report exhibited a good correlation and satisfactory agreement between experiment and theory.

ARTICLE HISTORY

Received: February 09, 2021
Revised: June 04, 2021
Accepted: June 21, 2021

DOI:
10.2174/1386207324666210816125738



Keywords: Chalcones, MAO-B, 2D-QSAR, 3D-QSAR, pharmacophore modeling, Q^2_{ext} .

1. INTRODUCTION

Designing a new class of inhibitors against Monoamine oxidases (MAOs) is a promising drug design strategy for various neurodegenerative disorders [1]. The enzyme is found in two isoforms, viz., MAO-A, and MAO-B, which metabolize amines in the brain and peripheral tissues [2]. During the oxidative deamination process of biogenic amines catalysed by MAOs, the hydrogen peroxide

by-product is formed which can further initiate the Reactive Oxygen Species (ROS) formation. The generation of these reactive free radicals promotes oxidative damage in the neuronal tissues [3]. Many of the clinical studies evidenced that elevated expression of MAO-B can manipulate neurotoxicity, which leads to Parkinson's Diseases (PD), Alzheimer's Disease (AD) etc., which are commonly known as neurodegenerative diseases [4]. The first line class of MAO-B inhibitors like deprenyl and rasagiline are irreversible in nature, making covalent bond interaction with flavin unit of MAO-B and propargyl group of drugs [5]. These suicidal inhibitors show poor pharmacokinetic profile, target disruption, and immunogenicity of enzyme-inhibitor adducts [6].

Some works have been performed using QSAR models and MAO-B. A description of ligand-based models to

*Address correspondence to these authors at the Department of Pharmaceutical Chemistry, Amrita School of Pharmacy, Amrita Vishwa Vidyapeetham, AIMS Health Sciences Campus, Kochi-682 041, India; Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900, João Pessoa, Brazil;
E-mails: bijomathew@aims.amrita.edu, bijovilaventgu@gmail.com (B. Mathew); mtscotti@gmail.com (M. T. Scotti)

[#]These authors contributed equally to this work.

5. MolPredictX: Online Biological Activity Predictions by Machine Learning Models.

Journal: Molecular Informatics

Impact Factor: 3.600

Check for updates

Research Article

www.molinf.com

doi.org/10.1002/minf.202200133

MolPredictX: Online Biological Activity Predictions by Machine Learning Models

Marcus Tullius Scotti,^[a] Chonny Herrera-Acevedo,^[a, b] Renata Priscila Barros de Menezes,^[a] Holli-Joi Martin,^[c] Eugene N. Muratov,^[c] Ávila Ítalo de Souza Silva,^[a] Emmanuella Faustino Albuquerque,^[a] Lucas Ferreira Calado,^[a] Ericsson Coy-Barrera,^[d] and Luciana Scotti.^[a]

Abstract: Here we report the development of MolPredictX, an innovative and freely accessible web interface for biological activity predictions of query molecules. MolPredictX utilizes in-house QSAR models to provide 27 qualitative predictions (active or inactive), and quantitative probabilities for bioactivity against parasitic (*Trypanosoma* and *Leishmania*), viral (Dengue, Sars-CoV and Hepatitis C), pathogenic yeast (*Candida albicans*), bacterial (*Salmonella enterica* and *Escherichia coli*), and Alzheimer disease enzymes. In this article, we introduce the methodology and usability of this webtool, highlighting its potential role in the development of new drugs against a variety of diseases. MolPredictX is undergoing continuous development and is freely available at <https://www.molpredictx.ufpb.br/>.

Keywords: MolPredictX · Online web-based tool · biological activity · Qualitative prediction · Drug development

1 Introduction

Quantitative Structure Activity Relationship (QSAR) modeling is a cheminformatics method that establishes a relationship between a molecule's biological activity and its structure.^[1–5] Studies by Hammett in 1937 and 1940^[6,7] initiated our understanding of molecular properties of organic compounds by establishing a linear relationship between varying substituent groups and the impact on biological properties. Subsequently, this relationship became known as the Hammett equation.

Decades later, Hansch and Fujita laid the foundations for QSAR/QSPR studies.^[8,9] In their pioneering study in 1962, Hansch and Fujita demonstrated that biological activity could be linearly correlated with different physicochemical parameters, related to hydrophobic, steric and electronic effects.^[8,9] Providing scientific support for the modeling of a biochemical property.

Over 60 years later, these models are still being developed^[8–10] using the principle that the molecular structure of a compound contains information about physical, chemical, and biological properties; in other words, similar chemical structures will have similar properties.^[5,10] Machine learning-based QSAR models are based on this same principle^[10] and have been extremely useful in to guide synthesis efforts and aid in early drug design.^[11,12] The use of these models also save time and decrease the costs associated with experimental procedures.^[10,13–20] However, like all computational methods, it is extremely important that QSAR models are properly validated before being used to interpret and predict the biological responses of compounds.^[21]

In recent years, several web servers have emerged to develop and apply Quantitative Structure Activity Relationship (QSAR) models using the Predictive Model Markup Language (PMML) format. Chembench, developed by Capuzzi, et al., in 2017^[22] is an integrated cheminformatics portal that automatically executes the complete QSAR process. Chembench uses five available machine learning algorithms (random forest (RF), support vector machine (SVM), k-nearest neighbors (kNN), genetic algorithm (GA) and simulated annealing (SA)) to perform ligand-based virtual screening. In the Chembench portal, it is also possible to use the QSAR MuDRA model, developed by Alves et al., 2018,^[23] which is conceptually related to the well-known kNN approach, but uses different types of chemical descriptors simultaneously for similarity assessment, providing a powerful alternative to consensus QSAR modeling.

[a] M. Tullius Scotti, C. Herrera-Acevedo, R. P. Barros de Menezes, Á. Ítalo de Souza Silva, E. Faustino Albuquerque, L. Ferreira Calado, L. Scotti.
Programa de Pós-Graduação de Produtos Naturais e Sintéticos Bioativos, Universidade Federal da Paraíba, 58051-900 João Pessoa-PB, Brazil
E-mail: mtsconfig@gmail.com

[b] C. Herrera-Acevedo
Department of Chemical Engineering, Universidad ECCI, Carrera 19 # 49-20, 111311, Bogotá D.C., Colombia

[c] H.-J. Martin, E. N. Muratov
Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA

[d] E. Coy-Barrera
Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia

Wiley Online Library

© 2022 Wiley-VCH GmbH

Mol. Inf. 2022, 41, 2200133

(1 of 9) 2200133

6. Selene-Ethylenelactamides and N-aryl-propanamides as Broad-Spectrum Leishmanicidal Agents-

Journal: Pathogens

Impact Factor: 3.700



Article

Selene-Ethylenelactamides and N-Aryl-Propanamides as Broad-Spectrum Leishmanicidal Agents

Natália Ferreira de Sousa ¹, Helivaldo Diógenes da Silva Souza ², Renata Priscila Barros de Menezes ¹, Francinara da Silva Alves ², Chonny Alexander Herrera Acevedo ¹, Thaís Amanda de Lima Nunes ³, Zoe L. Sessions ⁴, Luciana Scotti ¹, Eugene N. Muratov ⁴, Francisco Jaime Bezerra Mendonça-Junior ¹, Klinger Antônio da Franca Rodrigues ³, Petronio Filgueiras de Athayde Filho ² and Marcus Tullius Scotti ^{1,*}

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil

² Post-Graduate Program in Chemistry, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil

³ Infectious Diseases Laboratory, Federal University of Paraíba, Av. São Sebastião, nº 2819-Nossa Sra. de Fátima, Paraíba 64202-020, PI, Brazil

⁴ Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA

* Correspondence: mtsconfig@ccae.ufpb.br; Tel.: +55-83-99869-0415



Citation: de Sousa, N.F.; da Silva Souza, H.D.; de Menezes, R.P.B.; da Silva Alves, F.; Acevedo, C.A.H.; de Lima Nunes, T.A.; Sessions, Z.L.; Scotti, L.; Muratov, E.N.; Mendonça-Junior, F.J.B.; et al. Selene-Ethylenelactamides and N-Aryl-Propanamides as Broad-Spectrum Leishmanicidal Agents. *Pathogens* **2023**, *12*, 136. <https://doi.org/10.3390/pathogens12010136>

Academic Editors: Edson Roberto da Silva and Simone Brogi

Received: 5 November 2022

Revised: 20 December 2022

Accepted: 26 December 2022

Published: 13 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The World Health Organization classifies *Leishmania* as one of the 17 “neglected diseases” that burden tropical and sub-tropical climate regions with over half a million diagnosed cases each year. Despite this, currently available anti-leishmania drugs have high toxicity and the potential to be made obsolete by parasite drug resistance. We chose to analyze organoselenides for leishmanicidal potential given the reduced toxicity inherent to selenium and the displayed biological activity of organoselenides against *Leishmania*. Thus, the biological activities of 77 selenoesters and their N-aryl-propanamide derivatives were predicted using robust in silico models of *Leishmania infantum*, *Leishmania amazonensis*, *Leishmania major*, and *Leishmania (Viannia) braziliensis*. The models identified 28 compounds with >60% probability of demonstrating leishmanicidal activity against *L. infantum*, and likewise, 26 for *L. amazonensis*, 25 for *L. braziliensis*, and 23 for *L. major*. The in silico prediction of ADMET properties suggests high rates of oral absorption and good bioavailability for these compounds. In the in silico toxicity evaluation, only seven compounds showed signs of toxicity in up to one or two parameters. The methodology was corroborated with the ensuing experimental validation, which evaluated the inhibition of the Promastigote form of the *Leishmania* species under study. The activity of the molecules was determined by the IC₅₀ value (μM); IC₅₀ values < 20 μM indicated better inhibition profiles. Sixteen compounds were synthesized and tested for their activity. Eight molecules presented IC₅₀ values < 20 μM for at least one of the *Leishmania* species under study, with compound NC34 presenting the strongest parasite inhibition profile. Furthermore, the methodology used was effective, as many of the compounds with the highest probability of activity were confirmed by the in vitro tests performed.

Keywords: leishmaniasis; N-aryl-propanamides; selene-ethylenelactamides; CADD; organic synthesis

1. Introduction

The World Health Organization classifies *Leishmania* as one of the 17 “neglected diseases” that burden primarily tropical and sub-tropical climate regions [1]. These neglected diseases are caused by parasitic agents and are considered endemic in low-income populations. To date, they have been identified in 149 countries and are responsible for anywhere between 500,000 and 1,000,000 cases annually [2].

This infectious pathology comes from the Trypanosomatidae *Leishmania* sp protozoa [3,4] being transmitted to humans through the bite of infected female sand flies [5]. Among the existing *Leishmania* species, *Leishmania infantum* is a flagellated protozoan that causes

7. Methyleugenol Has an Antidepressant Effect in a Neuroendocrine Model: *In Silico* and *In Vivo* Evidence.

Journal: Pharmaceuticals

Impact Factor: 4.600



pharmaceuticals



Article

Methyleugenol Has an Antidepressant Effect in a Neuroendocrine Model: *In Silico* and *In Vivo* Evidence

Mayara Cecile Nascimento Oliveira ¹, Ikla Lima Cavalcante ¹, Alana Natalícia de Araújo ¹, Aline Matilde Ferreira dos Santos ¹, Renata Priscila Barros de Menezes ², Chonny Herrera-Acevedo ², Natália Ferreira de Sousa ², Jailane de Souza Aquino ³, José Maria Barbosa-Filho ⁴, Ricardo Dias de Castro ¹, Reinaldo Nóbrega Almeida ¹, Luciana Scotti ², Marcus Tullius Scotti ² and Mirian Graciela Da Silva Stiebbe Salvadori ^{1,*}

¹ Laboratory of Psychopharmacology, Institute for Research in Drugs and Medicines, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil; eualineferrer@gmail.com (A.M.F.d.S.); rcastro@ccs.ufpb.br (R.D.d.C.)

² Laboratory of Cheminformatics, Institute for Research in Drugs and Medicines, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil

³ Laboratory of Experimental Nutrition, Department of Nutrition, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil

⁴ Department of Pharmaceutical Sciences, Institute for Research in Drugs and Medicines, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil

* Correspondence: mirian.salvadori@gmail.com; Tel.: +55-83987430004



Citation: Oliveira, M.C.N.; Cavalcante, I.L.; de Araújo, A.N.; Ferreira dos Santos, A.M.; de Menezes, R.P.B.; Herrera-Acevedo, C.; Ferreira de Sousa, N.; de Souza Aquino, J.; Barbosa-Filho, J.M.; de Castro, R.D.; et al. Methyleugenol Has an Antidepressant Effect in a Neuroendocrine Model: *In Silico* and *In Vivo* Evidence. *Pharmaceuticals* **2023**, *16*, 1408. <https://doi.org/10.3390/ph16101408>

Academic Editors: Angel Josabad Alonso-Castro and Hye Won Lee

Received: 1 July 2023

Revised: 18 August 2023

Accepted: 21 August 2023

Published: 4 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Major depressive disorder is a severe mood disorder characterized by different emotions and feelings. This study investigated the antidepressant activity of the phenylpropanoid methyleugenol (ME) in adult female mice exposed to a stress model induced by dexamethasone. The animals were randomly divided into groups containing eight animals and were pre-administered with dexamethasone (64 µg/kg subcutaneously). After 165 and 180 min, they were treated with ME (25, 50 and 100 mg/kg intraperitoneally) or imipramine (10 mg/kg intraperitoneally) after 45 min and 30 min, respectively; they were then submitted to tests which were filmed. The videos were analyzed blindly. In the tail suspension test, ME (50 mg/kg) increased latency and reduced immobility time. In the splash test, ME (50 mg/kg) decreased grooming latency and increased grooming time. In the open field, there was no statistical difference for the ME groups regarding the number of crosses, and ME (50 mg/kg) increased the number of rearing and time spent in the center. Regarding *in silico* studies, ME interacted with dopaminergic D1 and α1 adrenergic pathway receptors and with tryptophan hydroxylase inhibitor. In the *in vivo* evaluation of the pathways of action, the antidepressant potential of ME (50 mg/kg) was reversed by SCH23390 (4 mg/kg intraperitoneally) dopaminergic D1 receptor, Prazosin (1 mg/kg intraperitoneally) α1 adrenergic receptor, and PCPA (4 mg/kg intraperitoneally) tryptophan hydroxylase inhibitor. Our findings indicate that ME did not alter with the locomotor activity of the animals and shows antidepressant activity in female mice with the participation of the D1, α1 and serotonergic systems.

Keywords: phenylpropanoid; dopaminergic; α1 adrenergic; serotonergic; *in vivo*; *in silico*

1. Introduction

Major depressive disorder (MDD) is a mood disturbance influenced by different emotions and feelings. Typically, MDD may feature feelings of sadness and helplessness, reduced or lost interest in daily activities (anhedonia), lack of motivation, depressed mood, irritability, and even suicidal ideation [1,2].

The World Health Organization (WHO) reported that in 2017 there were about 322 million people with depression worldwide (World Health Organization, 2017). The first year of the COVID-19 pandemic intensified the prevalence by 25% and the records show that,

8. Oxazolidinones from *Neocalyptrocalyx longifolium* Inhibit MsrA Protein in Methicillin Resistant *Staphylococcus aureus*.

Journal: Revista Brasileira de Farmacologia.

Impact Factor: 1.600

Revista Brasileira de Farmacognosia
<https://doi.org/10.1007/s43450-023-00422-6>

SHORT COMMUNICATION



Oxazolidinones from *Neocalyptrocalyx longifolium* Inhibit MsrA Protein in Methicillin Resistant *Staphylococcus aureus*

Thalisson Amorim de Souza¹ · Joanda Paola Raimundo Silva¹ · Damara Freitas Rodrigues² · Chonny Herrera-Acevedo¹ · Renata Priscila Barros de Menezes¹ · Nathalie Hellen Borges² · José Iranildo Miranda de Melo³ · José Pinto de Siqueira-Júnior² · Marcus Tullius Scotti¹ · Lucas Silva Abreu⁴ · Josean Fachine Tavares¹ · Marcelo Sobral da Silva¹

Received: 5 April 2023 / Accepted: 16 June 2023
© The Author(s) under exclusive licence to Sociedade Brasileira de Farmacognosia 2023

Abstract

The emergence of antimicrobial resistance has generated global concerns regarding many pathogenic microorganisms, such as methicillin-resistant *Staphylococcus aureus*. The inhibition of microbial molecular targets by natural products has led to the discovery of new paths capable of reverting resistance to classical antimicrobial agents. *Neocalyptrocalyx longifolium* (Mart.) Cornejo & Iltis, Capparaceae, is a Brazilian medicinal plant indicated for the treatment of skin and respiratory tract bacterial infections. Nevertheless, few studies have investigated its chemical composition. In view of the current development of pathogenic microorganism resistance, the isolation and identification of efflux pumps inhibitors from the roots of *N. longifolium* is described herein. In addition, the elements that contribute to substrate binding and inhibition of the MsrA protein, an ABC-type transporter, were analyzed based on *in silico* experiments. Five substances were isolated and characterized by NMR and HRMS. Four of them exhibited interesting structural features, composed of 1,3-oxazolidine-2-thione and 1,3-oxazolidine-2-one cores. 5-Methyl-5-ethyl-oxazolidine-2-one, an undescribed natural product, inhibited the activity of the MsrA transporter and, therefore, the potency of erythromycin was increased. Docking analysis revealed specific hydrogen interactions for this inhibitor at the MsrA ATP binding site.

Keywords Antimicrobial resistance · Caatinga · Erythromycin · Modulatory activity · Multidrug resistance · Oxazolidinone

Introduction

Methicillin-resistant *Staphylococcus aureus* (MRSA) causes severe infections in hospitals and communities. In recent decades, it has become a public health concern worldwide. Among MRSA's mechanisms of resistance, efflux pumps (EFPs) play a pivotal role because they act as the first-line defense mechanism against antimicrobials (Chalmers and

Wylam 2020) by decreasing the intracellular concentration of therapeutic antibiotics, clinical drugs, and other xenobiotics (Huang et al. 2022).

Efflux pump inhibitors, the so-called “magic bullets”, have become an important frontline to circumvent bacterial resistance. Natural products have shown substantial potential against EFPs in Gram-positive bacteria. Regarding MRSA, approximately 30 EFPs have been identified (Schindler and Kaatz 2016), with many studies focusing on the inhibition of major facilitator superfamily (MFS) transporters, such as NorA (Lira-Ricárdez and Pereda-Miranda 2020). However, few papers have reported natural substances that are active against the MsrA protein (Pinheiro et al. 2022). First described in the 1990s, MsrA is a transmembrane protein with two ATP binding motifs and 488 amino acids. This EFP is responsible for resistance to macrolides and streptogramins, both of which are important agents for treating Gram-positive bacterial infections in humans (Svetlov et al. 2021).

Oxazolidinethione and oxazolidinone are 5-member heterocyclic rings widely used in organic synthesis. The

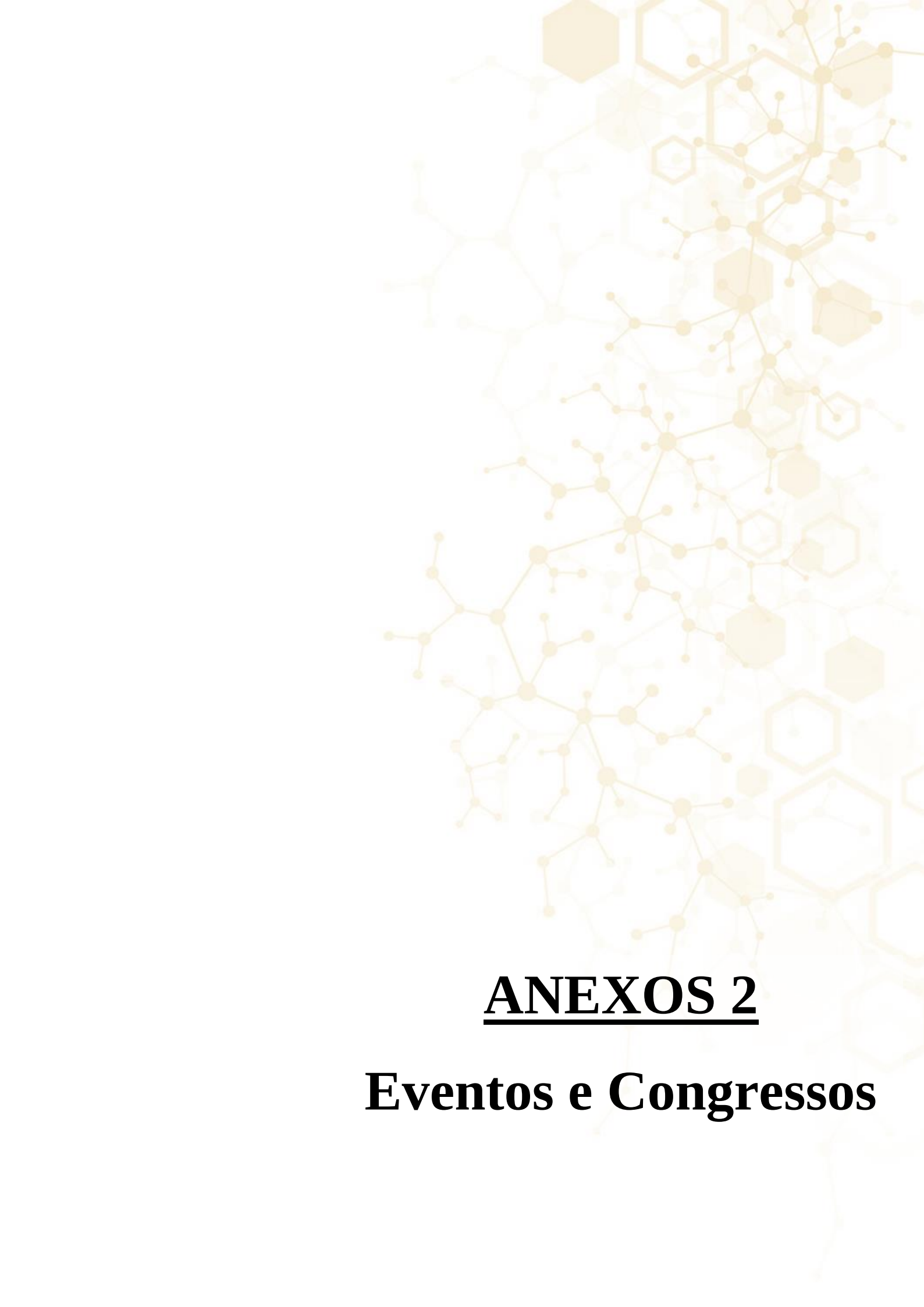
✉ Josean Fachine Tavares
josean@lft.ufpb.br

¹ Instituto de Pesquisa em Fármacos e Medicamentos, Universidade Federal da Paraíba, João Pessoa, PB, Brazil

² Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, PB, Brazil

³ Departamento de Biologia, Universidade Estadual da Paraíba, Campina Grande, PB, Brazil

⁴ Departamento de Química Orgânica, Universidade Federal Fluminense, Rio de Janeiro, RJ, Brazil



ANEXOS 2

Eventos e Congressos

1. 2nd Brazil France Symposium on Medicinal Chemistry

João Pessoa (Brazil) – outubro 2023

Modalidade: Poster



2. XIII Simpósio Brasileiro de Farmacognosia

Evento Online, Brasil – outubro 2021

Modalidade: Poster



3. 34° Congreso Latinoamericano de Química CLAQ 2020

Cartagena (Colômbia) – outubro 2021

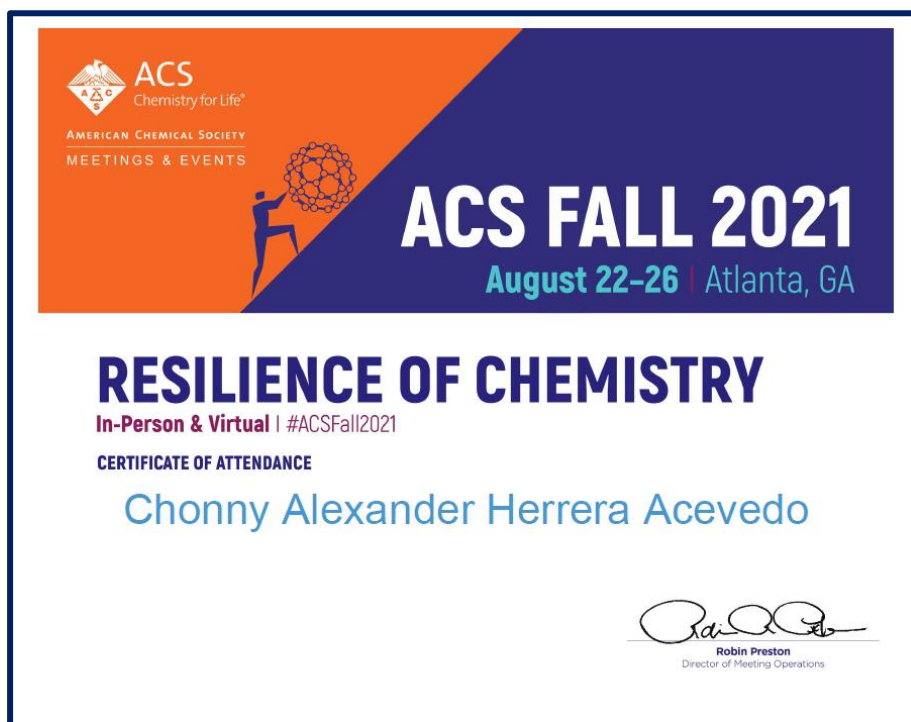
Modalidade: Poster



4. American Chemical Society (ACS) Fall 2021

Evento Online, USA – agosto 2021

Modalidade: Apresentação oral



1. ResNet NPND 10th anniversary 2011-2021

Evento Online, Alemanha – agosto 2021

Modalidade: Apresentação oral



Brazil – Colombia cooperation in cheminformatics studies against leishmaniasis: looking for terpenoid-based hits.

Chonny Herrera-Acevedo^{1,2*}, Marcus Tullius Scotti¹ and Ericsson Coy-Barrera²

¹ Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa, PB 58051-900, Brazil.

² Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia.

*presenting author; chonny622@gmail.com

Within the framework of a regional cooperation (Brazil-Colombia) between ResNet NPND members, a series of studies in cheminformatics have been performed, looking for terpenoid-based hits against leishmaniasis, a group of neglected tropical diseases that affect more than one million people worldwide [1]. Initially, an *in-house* database of 360 kauranes (tetracyclic diterpenes) was generated, and a combined ligand- and structure-based virtual screening (VS) approach was performed to select potential inhibitors of *Leishmania major* (Lm) pteridine reductase I (LmPTR1). For the ligand-based virtual screening, a machine learning classification model was built from the ChEMBL dataset (657 structures) which was classified as either active or inactive (binary classification). Sensitivity values of 78.1% and 82.6 % and specificity values of 72.7% and 73.7%, were obtained for the cross-validation and test sets, respectively. Only 7 from 360 structures were classified as active (ligand-based probability value [LB] ≥ 0.5), with structures **135** (2 β -hydroxy-menth-6-en-5 β -yl *ent*-kaurenoate) and **302** (3 α -cinnamoyloxy-*ent*-kaur-16-en-19-oic acid) representing two of the best-ranked kauranes, with LB values of 0.57 and 0.54, respectively. These two kauranes were employed to verify the validity of the VS approach through LmPTR1 enzyme inhibition assay. The half-maximal inhibitory concentration (IC₅₀) values of selected bioactive compounds were below 10 μ M, as predicted in the classification model. A compound structurally related to **302**, 3 α -*p*-coumaroyloxy-*ent*-kaur-16-en-19-oic acid (**302a**), was also synthesized and showed the highest activity against LmPTR1. Finally, molecular docking calculations and molecular dynamics simulations were performed for the VS-selected, most-active kauranes within the active sites of PTR1 hybrid models, generated from three *Leishmania* species that are known to cause cutaneous leishmaniasis in the new world (i.e., *L. braziliensis*, *L. panamensis*, and *L. amazonensis*) [2,3] to explore the targeting potential of these kauranes to other species-dependent variants of this enzyme.

[1] López-Arencibia, A.; Bethencourt-Estrella, C. J.; Freijo, M. B.; Reyes-Batlle, M.; Sifaoui, I.; San Nicolás-Hernández, D.; et al. New phenalenone analogues with improved activity against *Leishmania* species. *Biomedicine & Pharmacotherapy* **2020**, *132*, 110814. doi: 10.1016/j.biopha.2020.110814

[2] Sánchez-Suárez, J.; Bernal, F. A.; Coy-Barrera, E., Colombian Contributions Fighting Leishmaniasis: A Systematic Review on Antileishmanials Combined with Chemoinformatics Analysis. *Molecules* **2020**, *25*, (23), 5704. doi: 10.3390/molecules25235704

[3] Anversa, L.; Tiburcio, M. G. S.; Richini-Pereira, V. B.; Ramirez, L. E., Human leishmaniasis in Brazil: a general review. *Revista da Associação Médica Brasileira* **2018**, *64*, (3), 281-289. doi: 10.1590/1806-9282.64.03.281

2. Ligand-based and Structure-based virtual screening for the discovery of natural larvicial against *Aedes aegypti*

Evento Online, MOL2NET – 2021

Modalidade: Resumo expandido

MOL2NET, 2021, 7, ISSN: 2624-5078

1



MOL2NET, International Conference Series on Multidisciplinary Sciences

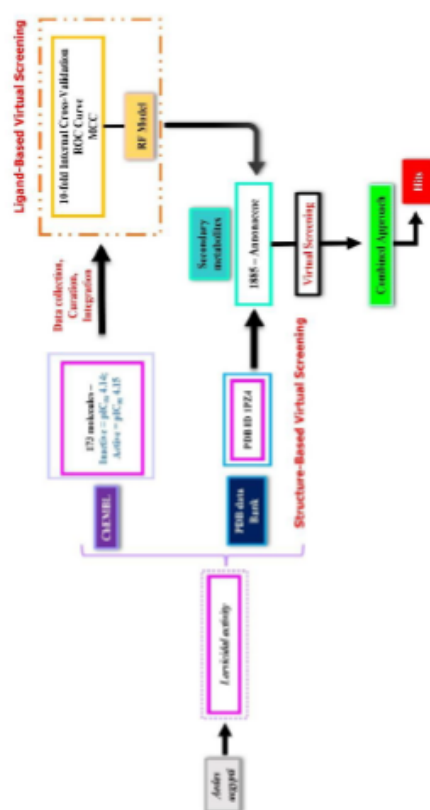
<http://sciforum.net/conference/mol2net-03>

Ligand-Based and Structure-based virtual screening for the discovery of natural larvicial against *Aedes aegypti*

<Renata Priscila Barros de Menezes> (renatabarros@luf.ufrpb.br)^a, <Chonny Herrera Acevedo> (chonny622@gmail.com)^a, <Luciana Scotti> (luciana.scotti@gmail.com)^a
<Marcus Tullius Scotti> (mtscotti@gmail.com)^a

^a < Post-Graduate Program in Natural Synthetic Bioactive Products, Federal university of Paraiba >

Graphical Abstract



Abstract. The *Aedes aegypti* mosquito belongs to the order Diptera and is one of the main vectors of transmission of etiological agents that cause several diseases. This mosquito can transmit diseases such as dengue, yellow fever, Zika, chikungunya, among others. The aim of this study was combining structure-based and ligand-based virtual screening (VS) techniques to select potentially larvicial active molecules against *Ae. aegypti* from in-house secondary metabolite dataset (SistematX). From the ChEMBL database, we selected a set of 161 chemical structures with larvicial activity against *Ae. aegypti* to create random forest models with an accuracy value higher than 82% for cross-validation and test sets. Afterward, the ligand-based virtual screen selected 38 secondary metabolites. In addition, a structure-based virtual screening was also performed for the 38 molecules selected. Finally, using consensus analyzes approach combining ligand-based and structure-based VS, five molecules were selected as potential larvicial against *Ae. aegypti*.

Keywords: Virtual Screening; Secondary Metabolites; Annonaceae; Larvicial activity; *Aedes aegypti*

3. Ligand-based and Structure-based virtual screening for the discovery of natural larvicidal against *Aedes aegypti*

Evento Online, MOL2NET – 2021

Modalidade: Resumo expandido

MOL2NET, 2021, 7, ISSN: 2624-5078
<https://mol2net-07.sciforum.net/>

1



**MOL2NET, International Conference
Series on Multidisciplinary Sciences**

FROM MOLECULES TO NETWORKS



UPV/EHU IKERBASQUE

**In silico Identification of Potential Sesquiterpene Lactones from
SistematX Database against *Schistosoma mansoni***

Chonny Herrera-Acevedo ^{a,b}, Renata Priscila Barros de Menezes ^a, Luciana Scotti ^a,
Ericsson Coy-Barrera ^b, Marcus Tullius Scotti ^{a*}.

^a Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba,
João Pessoa, PB 58051-900, Brazil.

^b Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar
Nueva Granada, Cajicá 250247, Colombia.

Graphical Abstract	Abstract. Schistosomiasis is an acute and chronic parasitic disease, caused by blood flukes (trematode worms) of the genus <i>Schistosoma</i>. For 2019, the World Health Organization estimated that close to 240 million people required preventive treatment against this disease, mainly poor communities without access to safe drinking water and adequate sanitation. Similarly, to others Neglected Tropical Diseases (NTDs) the treatments against this disease are limited, and new chemotherapies are necessary to the control
----------------------------------	--