

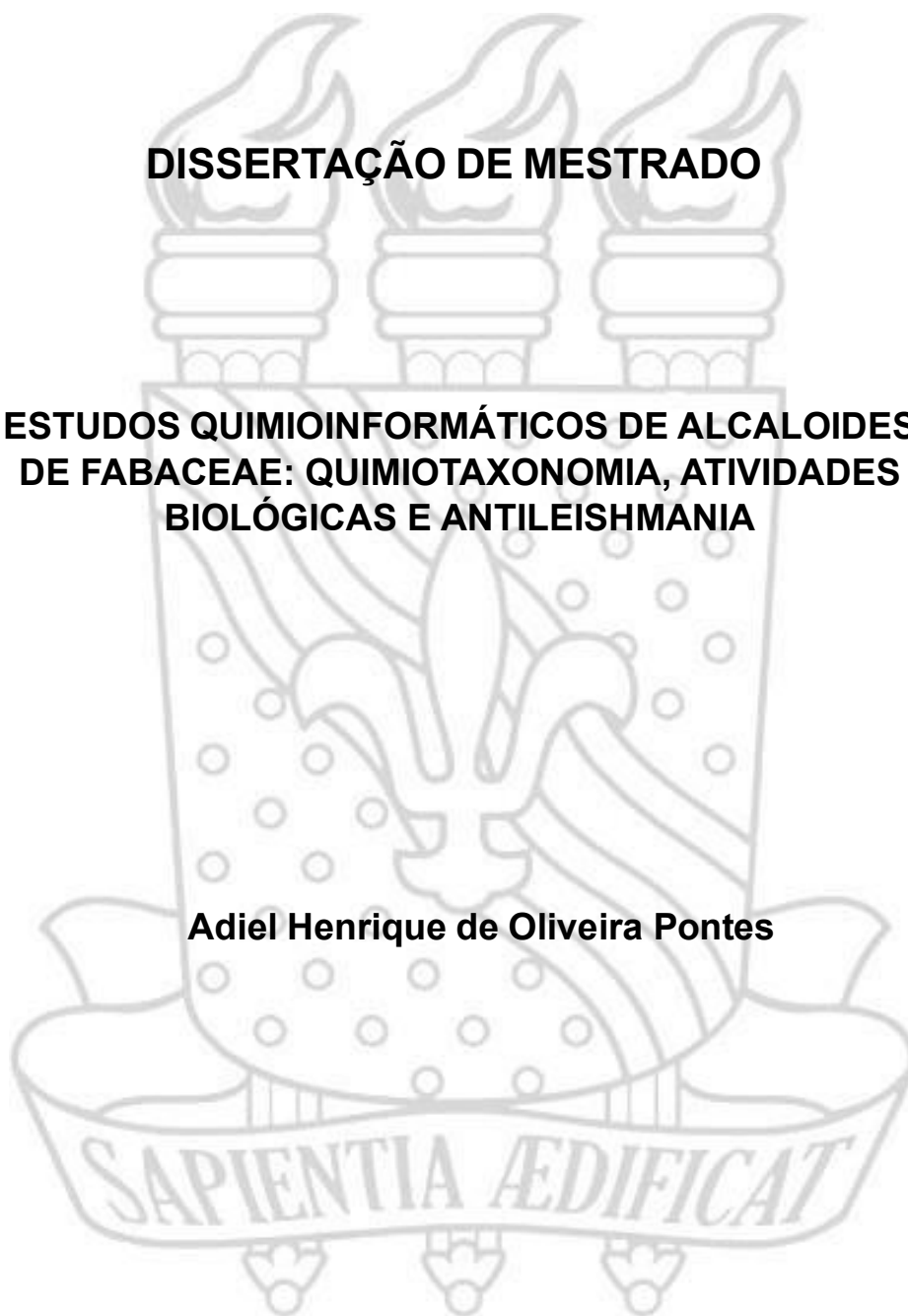


UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

ESTUDOS QUIMIOINFORMÁTICOS DE ALCALOIDES
DE FABACEAE: QUIMIOTAXONOMIA, ATIVIDADES
BIOLÓGICAS E ANTILEISHMANIA

Adiel Henrique de Oliveira Pontes



João Pessoa – PB

2025



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**ESTUDOS QUIMIOINFORMÁTICOS DE ALCALOIDES DE
FABACEAE: QUIMIOTAXONOMIA, ATIVIDADES BIOLÓGICAS E
ANTILEISHAMANIA**

**Adiel Henrique de Oliveira
Pontes***

Dissertação apresentada como
requisito para obtenção do título de
Mestre em Química pela
Universidade Federal da Paraíba.

Orientador: Prof. Dr. Marcus Tullius Scotti

*Bolsista CAPES

João Pessoa – PB

2025

Catálogo na publicação
Seção de Catalogação e Classificação

P814e Pontes, Adiel Henrique de Oliveira.

Estudos quimioinformáticos de alcaloides de fabaceae
: quimiotaxonomia, atividades biológicas e
antileishmania / Adiel Henrique de Oliveira Pontes. -
João Pessoa, 2025.
90 f. : il.

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

1. Alcaloides. 2. Quimiotaxonomia. 3. Metabolitos
secundários. I. Scotti, Marcus Tullius. II. Título.

UFPB/BC

CDU 547.94(043)

À Deus, à minha família, aos docentes e pesquisadores que compõem a comunidade científica, e aos meus passados, presentes e futuros alunos.

Dedico!

“A imaginação é mais importante que o conhecimento”

Albert Einstein

AGRADECIMENTOS

Agradeço a Deus por todo sustento e renovação concedida durante todo esse período e ao longo da minha vida.

Agradeço a CAPES pela bolsa concedida e a estrutura disponibilizada pela UFPB, em especial a do laboratório de quimioinformática.

Agradeço ao professor Marcus Tullius Scotti por todos os ensinamentos transmitidos ao longo desses dois anos, também sou grato pela relação de amizade que construímos e espero levava para toda a vida.

Agradeço a minha família por todo apoio e auxílio que sempre fez toda diferença, assim como aos meus colegas de laboratório (Alan, Túlio, Thais, Renata, Chonny, Natália e Luís) que sempre foram responsáveis por criar um ambiente agradável de aprendizado e crescimento.

Agradeço aos professores de todo o Departamento de Química da UFPB, em especial ao professor Wallace Duarte Fragoso e à professora Antônia Lúcia de Souza por todo apoio concedido nessa fase final.

Por fim, agradeço de forma geral a todos os amigos e colegas de profissão que fizeram e fazem parte da minha trajetória ao longo de todos esses anos.

RESUMO

Título: Estudos quimiinformáticos de alcaloides de Fabaceae: quimiotaxonomia, atividades biológicas e antileishmania

A quimiotaxonomia, método de classificação de organismos baseado em seus componentes químicos, tem se mostrado uma ferramenta valiosa na taxonomia vegetal. Essa estratégia, aliada à análise de produtos naturais, pode revelar informações sobre as relações evolutivas e adaptações ecológicas das espécies vegetais. Na família Fabaceae (Leguminosae), conhecida por sua capacidade de fixação de nitrogênio e diversidade química, a quimiotaxonomia tem sido particularmente útil para esclarecer relações taxonômicas. Alcaloides, um grupo importante de produtos naturais encontrados em Fabaceae, frequentemente exibem diversas atividades biológicas, incluindo propriedades anticancerígenas. Mapas auto-organizáveis (SOMs) são uma técnica de análise poderosa para visualizar e classificar conjuntos de dados complexos. Ao aplicar SOMs à análise de perfis de alcaloides em Fabaceae, podemos obter insights valiosos sobre a diversidade química dessa família e identificar potenciais marcadores quimiotaxonômicos. Essa abordagem pode auxiliar na descoberta de novos compostos bioativos e aprimorar nossa compreensão da história evolutiva de Fabaceae. Neste estudo, a aplicação de SOMs demonstrou resultados satisfatórios na distinção quimiotaxonômica de dois gêneros da família Fabaceae, *Sophora* e *Erythrina*, com taxas de acerto superiores a 90% tanto no geral quanto para cada gênero isoladamente. Em suma, é possível concluir que estruturas químicas específicas podem estar relacionadas a diferentes biossínteses dos metabólitos secundários produzidos por espécies de plantas desses dois gêneros, assim como atividades biológicas específicas de cada gênero. Além disso, através da combinação do docking molecular com a predição de propriedades farmacocinéticas foi possível selecionar três alcaloides do banco inicial que se apresentam como potenciais fármacos contra leishmaniose.

Palavras-chave: SOMs; quimiotaxonomia; metabólitos secundários; alcaloides; docking molecular; ADMET.

ABSTRACT

Title: Chemoinformatic studies of Fabaceae alkaloids: chemotaxonomy, biological activities, and anti-leishmanial potential

Chemotaxonomy, a method of classifying organisms based on their chemical components, has proven to be a valuable tool in plant taxonomy. This strategy, combined with natural product analysis, can reveal information about the evolutionary relationships and ecological adaptations of plant species. In the Fabaceae (Leguminosae) family, known for its nitrogen-fixing ability and chemical diversity, chemotaxonomy has been particularly useful in clarifying taxonomic relationships. Alkaloids, an important group of natural products found in Fabaceae, often exhibit various biological activities, including anticancer properties. Self-organizing maps (SOMs) are a powerful analysis technique for visualizing and classifying complex datasets. By applying SOMs to the analysis of alkaloid profiles in Fabaceae, we can gain valuable insights into the chemical diversity of this family and identify potential chemotaxonomic markers. This approach can assist in the discovery of new bioactive compounds and enhance our understanding of the evolutionary history of Fabaceae. In this study, the application of SOMs demonstrated satisfactory results in the chemotaxonomic distinction of two genera of the Fabaceae family, *Sophora* and *Erythrina*, with accuracy rates exceeding 90% both overall and for each genus individually. In summary, it is possible to conclude that specific chemical structures may be related to different biosyntheses of secondary metabolites produced by plant species of these two genera, as well as specific biological activities of each genus. Furthermore, through the combination of molecular docking with the prediction of pharmacokinetic properties, it was possible to select three alkaloids from the initial database that present themselves as potential drugs against leishmaniasis.

Keywords: SOMs; chemotaxonomy; secondary metabolites; alkaloids; docking molecular; ADMET;

LISTA DE ABREVIATURAS

RMSD Desvio Médio Quadrático da Raiz (do inglês, Root Mean Square Deviation)

SOMs Mapas Auto-Organizáveis (do inglês, Self-Organizing Maps)

ACP Análise de Componente Principal

LogP Coeficiente de Partição Octanol-Água

tPAS Área de Superfície Topológica (do inglês, Topological Polar Surface Area)

LISTA DE FIGURAS

Figura 1. Biossíntese de alcaloides quinolizidínicos e piperidínicos de <i>Sophora</i> -----	22
Figura 2. Biossíntese de alcaloides isoquinolínicos de <i>Erythrina</i> -----	23
Figura 3. Uma visão geral da filogenia de Fabaceae -----	26
Figura 4. Visão geral da formulação de modelos de aprendizado de máquina -----	31
Figura 5. Ilustração da divisão de dados (subconjuntos de treino e teste) realizada para a validação cruzada do tipo 5-fold -----	32
Figura 6. Comparação das validações interna e externa -----	33
Figura 7. Mapa auto-organizável obtido pela classificação dos gêneros <i>Erythrina</i> e <i>Sophora</i> : (a) SOM dos descritores moleculares; (b) U-Matrix; (c) Projeção PCA do SOM -----	56
Figura 8. Mapa auto-organizável obtido pela classificação dos gêneros <i>Erythrina</i> e <i>Sophora</i> : (a) Descritor molecular nCXr; (b) Descritor molecular nRC=N; (c) Descritor molecular nR10; (d) Posição dos compostos 1-6 no SOM -----	57
Figura 9. Alcaloides que exibem a diferença mostrada pelo descritor nCXr no SOM -----	58
Figura 10. Alcaloides que exibem a diferença mostrada pelo descritor nRC=N no SOM ----	59
Figura 11. Alcaloides que exibem a diferença mostrada pelo descritor nR10 no SOM -----	59
Figura 12. Interações com os resíduos de aminoácidos da ornitina decarboxilase -----	62
Figura 13. Interações com os resíduos de aminoácidos 14-alfa-desmetilase -----	64
Figura 14. Interações com os resíduos de aminoácidos da PnC1 -----	66
Figura 15. Interações com os resíduos de aminoácidos da piridoxal quinase -----	68
Figura 16. Alcaloides com melhores resultados farmacocinéticos nos parâmetros analisados -----	72

LISTA DE TABELAS

Tabela 1. Conjunto de descritores gerados para as moléculas do banco de dados -----	45
Tabela 2. Alvos e ligantes usados no docking molecular -----	50
Tabela 3. Distribuição das moléculas encontradas, classificadas por gênero, incluindo dados botânicos e ocorrências químicas -----	53
Tabela 4. Resultados do mapa auto-organizável com os valores das ocorrências e o número de acertos corretos para os gêneros <i>Erythrina</i> e <i>Sophora</i> da família Fabaceae -----	54
Tabela 5. Resumo dos resultados de correspondência de treinamento e teste em porcentagem de <i>Erythrina</i> e <i>Sophora</i> -----	55
Tabela 6. Cinco melhores moléculas obtidas no docking com a ornitina descarboxilase ----	61
Tabela 7. Cinco melhores moléculas obtidas no docking com a 14-alfa-desmetilase -----	63
Tabela 8. Cinco melhores moléculas obtidas no docking com a PnC1 -----	65
Tabela 9. Cinco melhores moléculas obtidas no docking com a piridoxal quinase -----	67
Tabela 10. Propriedades de lipofilicidade e área de superfície polar topológica para as moléculas analisadas, destacando aquelas com bons resultados em um parâmetro (negrito) ou ambos (vermelho) -----	70
Tabela 11. Druglikeness das moléculas analisadas, destacando aquelas com no máximo uma violação em todas as regras -----	71

SUMÁRIO

1. INTRODUÇÃO	13
2. FUNDAMENTAÇÃO TEÓRICA	16
2.1 Família Fabaceae e seus alcaloides	17
2.2 Filogenia e biossíntese das espécies de Fabaceae	17
2.3 Quimiotaxonomia e mapas auto-organizáveis	24
2.4 Redes neurais e mapas auto-organizáveis	27
2.5 Validação de modelos de aprendizado de máquina e análise de componente principal	30
2.6 Doenças Tropicais Negligenciada: Leishmaniose	35
2.7 Farmacocinética e Propriedades ADMET	36
2.8 Docking Molecular	37
3. OBJETIVOS	42
3.1 Objetivo Geral	43
3.2 Objetivos Específicos	43
4. METODOLOGIA	44
4.1 Banco de dados dos alcaloides	45
4.2 Descritores Moleculares	45
4.3 Mapas auto-organizáveis e 5-fold	49
4.4 Docking Molecular e Propriedades ADMET	49
5. RESULTADOS E DISCUSSÕES	52
5.1 Banco de Dados	53
5.2 Mapas auto-organizáveis e descritores moleculares aplicados à quimiotaxonomia de alcaloides de Fabaceae	53
5.3 Biossíntese de alcaloides de <i>Erythrina</i> e <i>Sophora</i>	60
5.4 Atividade biológica dos alcaloides de <i>Erythrina</i> e <i>Sophora</i>	60
5.5 Docking Molecular e Propriedades ADMET	61
6. CONCLUSÕES	74
7. REFERÊNCIAS	76
APÊNDICES	85

CAPÍTULO 1: INTRODUÇÃO

1. INTRODUÇÃO

A classificação de organismos com base em seus constituintes químicos provou ser uma ferramenta valiosa no estudo da taxonomia de plantas; nesse sentido, a quimiotaxonomia é um método de classificação sistemática que emprega a análise de compostos químicos presentes em organismos para determinar suas afinidades taxonômicas [1]. Essa abordagem, combinada com a análise de produtos naturais, pode fornecer insights sobre as relações evolutivas e adaptações ecológicas de espécies de plantas [16,17]. No caso da família Fabaceae (Leguminosae), um grupo diversificado de plantas conhecido por suas capacidades de fixação de nitrogênio e rica diversidade química, a quimiotaxonomia tem sido particularmente útil para elucidar relações taxonômicas [6].

Os gêneros *Erythrina* e *Sophora* são renomados por sua abundância de alcaloides, compostos nitrogenados que exibem uma gama diversificada de propriedades farmacológicas [8, 7]. Esses alcaloides possuem estruturas químicas complexas e variadas, conferindo a essas plantas um potencial significativo para o desenvolvimento de novos produtos farmacêuticos. Estudos indicaram que os alcaloides presentes em *Erythrina* e *Sophora* podem apresentar atividades biológicas intrigantes, como propriedades analgésicas, anti-inflamatórias, antimicrobianas e até anticancerígenas [9,15]. A diversidade estrutural desses compostos, juntamente com sua bioatividade, torna esses gêneros botânicos alvos promissores para a pesquisa de produtos naturais.

Em termos de filogenia, tanto *Erythrina* quanto *Sophora* estão situados em clados bem definidos dentro da família Fabaceae. Estudos filogenéticos moleculares, exemplificados pelo trabalho de Zhao *et al.* [6], avançaram significativamente nossa compreensão das relações evolutivas entre esses gêneros e outros membros da família, revelando padrões de especiação e divergência ao longo de sua história evolutiva. Os posicionamentos filogenéticos precisos de *Erythrina* e *Sophora* são cruciais para compreender a evolução de seus atributos morfológicos, fisiológicos e bioquímicos, bem como para tirar inferências sobre a origem e diversificação dos alcaloides encontrados nessas plantas [18,19].

Nesse contexto, os mapas auto-organizáveis (do inglês, SOMs) emergem como uma ferramenta indispensável na quimiotaxonomia [20]. Ao analisar os perfis químicos

complexos de diferentes espécies, especialmente metabólitos secundários como alcaloides, os SOMs são capazes de gerar representações visuais que revelam intuitivamente os padrões e relações entre as amostras [20, 21]. Essa capacidade única permite a identificação de espécies quimicamente semelhantes, mesmo quando pertencem a gêneros ou famílias taxonomicamente distantes, contribuindo para a construção de classificações mais precisas e robustas [21]. Adicionalmente, os SOMs podem ser usados para identificar marcadores químicos específicos que caracterizam certos grupos taxonômicos, facilitando a descoberta de novos taxóons e aprofundando a compreensão da evolução química das plantas ao longo do tempo [22].

Nessa perspectiva, os SOMs mostram grande potencial para a identificação de padrões químicos nos alcaloides de *Erythrina* e *Sophora*. Ao analisar os perfis químicos complexos dessas plantas, os SOMs podem revelar agrupamentos de espécies com características químicas semelhantes, mesmo que pertençam a diferentes táxons [20]. Essa capacidade é particularmente útil para a identificação de marcadores químicos específicos associados a certos grupos taxonômicos, o que pode auxiliar na descoberta de novas espécies e na compreensão da evolução química desses gêneros [21, 22].

Compreendendo que essa família de plantas (Fabaceae) apresenta diversos alcaloides com importantes atividades biológicas, inclusive atividade antiparasitária como descrito por Rambo et al. [7] e Wang et al. [8], é possível entender que alcaloides dessa família, por consequência, possuem, no geral, potenciais atividades antiparasitárias [7, 8]. Nesse contexto, a aplicação de técnicas de triagem virtual para selecionar alcaloides com esse potencial, se mostra como uma ótima opção, inclusive no âmbito do combate as doenças tropicais negligenciadas (do inglês, NTDs), como a leishmaniose [58, 64]. Técnicas como simulações de docking molecular e predição de propriedades ADMET são bastante promissoras, quando aplicadas em conjunto, para a identificação de moléculas com potenciais atividades antiparasitárias, e com propriedades farmacocinéticas favoráveis, como baixa toxicidade, por exemplo [56, 84, 85].

A aplicação de técnicas *in silico* se apresenta como uma ótima opção no desenvolvimento de fármacos para as NTDs, visto que a falta de interesse das empresas farmacêuticas em desenvolver novos medicamentos, devido ao baixo retorno financeiro, é um dos principais obstáculos para controlar e eliminar essas doenças, além disso, os tratamentos atuais são frequentemente ineficazes, tóxicos e

de longa duração [60, 61]. Sendo assim, a utilização de técnicas *in silico* se apresenta como uma possibilidade bastante promissora, afinal essas possibilitam reduções nos custos de produção, redução no tempo de desenvolvimento, redução na matéria-prima utilizada e nos resíduos processuais relacionados ao desenvolvimento de um novo fármaco, em comparação com métodos tradicionais [68].

No presente estudo, um total de 374 alcaloides encontrados em 455 ocorrências dos gêneros *Erythrina* e *Sophora* foram analisados usando SOMs. Além disso, o banco por completo, 650 alcaloides em 747 ocorrências botânicas, foi triado, através do docking molecular e da predição das propriedades ADMET, com o intuito de selecionar moléculas com alto potencial leishmanicida e com propriedades farmacocinéticas favoráveis. Os dois principais objetivos da pesquisa foram identificar padrões químicos que pudessem diferenciar as moléculas presentes nos dois gêneros, visando estabelecer um sistema de classificação robusto, e selecionar moléculas que possuam potenciais propriedades para serem novos fármacos para leishmaniose.

CAPÍTULO 2: FUNDAMENTAÇÃO TEÓRICA

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Família Fabaceae e seus alcaloides

A família Fabaceae, também conhecida como Leguminosae, é um grupo de plantas de destaque nos ecossistemas do planeta. Com uma vasta diversidade, contabiliza-se cerca de 765 gêneros e 19.500 espécies, conforme dados do Grupo de Trabalho de Filogenia de Leguminosas [24]. Essa expressiva representação a coloca como a terceira maior família de angiospermas, ficando atrás apenas da Asteraceae e da Orchidaceae. As Fabaceae são encontradas em diversos habitats, desde os trópicos até as regiões temperadas, e desempenham papéis importantes na dinâmica ecológica desses locais [25]. Além disso, possuem grande relevância econômica, já que muitas espécies são utilizadas como alimento, fonte de madeira e outros produtos de valor. Em outras palavras, as leguminosas são um grupo de plantas muito importante e abundante, com grande impacto tanto na natureza quanto na economia.

As leguminosas, pertencentes à família Fabaceae, destacam-se por sua capacidade de fixar o nitrogênio atmosférico através de uma relação simbiótica com bactérias do gênero *Rhizobium*. Essa parceria ocorre nos nódulos radiculares, estruturas especializadas nas raízes das plantas, onde as bactérias transformam o nitrogênio gasoso em amônia, uma forma utilizável pelas plantas. Essa habilidade confere às leguminosas um grande sucesso ecológico, permitindo que prosperem em diversos ambientes, inclusive solos pobres em nitrogênio [26]. Além disso, a fixação biológica de nitrogênio pelas leguminosas desempenha um papel crucial no ciclo global desse elemento, contribuindo para a fertilidade do solo e influenciando a produção agrícola. Essa capacidade de fixar nitrogênio não é exclusiva das leguminosas, sendo compartilhada por membros de outras oito famílias de plantas, que também estabelecem parcerias com microrganismos fixadores de nitrogênio [27]. Essa característica demonstra a importância da fixação biológica de nitrogênio para a manutenção da vida vegetal em nosso planeta.

Com os grandes avanços obtidos nas áreas da fitoquímica e da química analítica, quimiosistemáticos notaram que certos compostos, denominados metabólitos secundários, apresentavam uma distribuição restrita no mundo vegetal, estando presentes apenas em grupos específicos de plantas, como famílias, tribos ou gêneros. Essa observação levou ao desenvolvimento da quimiotaxonomia, área que

utiliza a composição química das plantas como ferramenta para auxiliar na sua classificação e compreensão das relações evolutivas [28, 29]. Inicialmente, os metabólitos secundários eram considerados como compostos sem função aparente, talvez apenas produtos residuais do metabolismo vegetal. Nessa visão, eles seriam marcadores neutros, que podem ser definidos como compostos químicos produzidos pelas plantas que não sofrem forte influência de fatores ambientais e, portanto, refletem mais fielmente a herança genética de um organismo [105, 106]. Contudo, essa perspectiva foi revisada à luz de novas descobertas, por exemplo, compreendeu-se que fatores abióticos, como a composição de micro e macronutrientes no solo, e bióticos, como a ação de herbívoros e parasitas, influenciam a orquestração das biossínteses responsáveis pela produção de óleos vegetais em cada espécie. Dessa forma, a produção de metabólitos secundários (moléculas que compõem os óleos essenciais) está diretamente relacionada às condições ambientais, fazendo com que as rotas sintéticas sejam entendidas como resultados de processos adaptativos [30, 105, 106]. Atualmente, é reconhecida a importância crucial dos metabólitos secundários para a sobrevivência e o sucesso ecológico das plantas. Estes desempenham papéis diversos, como defesa contra herbívoros e patógenos, atração de polinizadores, proteção contra radiação UV e estresse ambiental, entre outros. Dessa forma, os metabólitos secundários não podem mais ser considerados marcadores neutros. A compreensão da sua função e do seu papel na interação das plantas com o ambiente é fundamental para a ecologia química e para diversas áreas da ciência, como a farmacologia e a biotecnologia [106].

As espécies de *Erythrina* estão incluídas na família Fabaceae, e se distribuem por regiões tropicais do mundo, estendendo-se por áreas temperadas quentes. Triagens fitoquímicas mostraram diversos grupos de compostos em espécies de *Erythrina*, como catequinas, esteroides, flavonoides, saponinas, triterpenos e xantonas [31, 32]. No entanto, alcaloides tetracíclicos são os principais marcadores químicos deste gênero, alguns desses compostos possuem diversas atividades biológicas, como erisodina que desempenha atividade ansiolítica [33]. Propriedades anticonvulsiantes, analgésicas, anti-inflamatórias e até mesmo anticâncer são observadas em outros alcaloides derivados de espécies pertencentes a este gênero de plantas [33].

Em relação as espécies do gênero *Sophora* (Fabaceae), pode-se afirmar, que esse gênero, possui mais de 15 espécies de plantas medicinais, como *Sophora*

flavescens, *Sophora alopecuroides* e *Sophora tonkinensis*. Alcaloides quinolizidínicos, uma das principais formas de alcaloides encontrados no gênero *Sophora*, foram isolados de diferentes partes medicinais da planta como flores, folhas, sementes ou raízes. Dentre esses compostos, pode-se destacar a matrina que possui propriedades terapêuticas relacionada a problemas cardiovasculares [34, 35]. Outras atividades biológicas importantes também foram observadas nos alcaloides desse gênero, e estão descritas por Wang *et al.* [36].

2.2 Filogenia e biossíntese dos gêneros *Sophora* e *Erythrina*

A filogenia, que se dedica a desvendar os caminhos da evolução entre grupos de organismos como espécies e populações, busca traçar um mapa de como esses grupos se conectam ao longo do tempo, como um quebra-cabeça que remonta à sua história evolutiva. Essa área de estudo é uma ferramenta poderosa em diversos campos da biologia, como a sistemática, que organiza e classifica os seres vivos, a ecologia, que estuda as interações entre os organismos e o meio ambiente, a biologia evolutiva, que investiga como as espécies se transformam ao longo do tempo, e até mesmo a medicina, que utiliza a filogenia para entender a evolução de doenças e desenvolver tratamentos. Através da filogenia, os cientistas podem classificar os organismos em grupos, como se estivessem montando uma árvore genealógica da vida, entender como a vida na Terra se diversificou ao longo de bilhões de anos, como um álbum de fotos da história natural, e estudar como características e comportamentos surgiram e se modificaram, como um livro aberto sobre a evolução. No entanto, apesar de ser amplamente aceito que a taxonomia biológica moderna se baseia em nosso conhecimento da história evolutiva, a aplicação da filogenia na taxonomia ainda não está completa [37].

Embora o princípio da descendência comum, que afirma que todos os seres vivos compartilham um ancestral comum, tenha ganhado destaque, ele não se tornou a base central da taxonomia. As taxonomias existentes, que já estavam em uso antes da filogenia, passaram a ser interpretadas à luz da evolução, como se fossem mapas antigos que ganharam novas legendas. Princípios evolutivos foram usados para justificar práticas taxonômicas estabelecidas, como se estivessem adaptando as regras do jogo para manter o placar. O princípio da descendência comum, apesar de sua importância, não se tornou o alicerce a partir do qual os princípios e métodos da

taxonomia são construídos. Em vez disso, ele foi utilizado para dar um verniz evolutivo a práticas preexistentes, como se estivessem pintando um quadro antigo com novas cores [38].

As plantas, por serem organismos fixos em seus habitats, desenvolveram uma estratégia para sobreviver e prosperar: a produção de metabólitos secundários. Estas substâncias químicas, além de atuarem na comunicação entre as plantas, são verdadeiros escudos moleculares, protegendo-as contra ameaças como herbívoros vorazes, plantas concorrentes que disputam recursos e patógenos oportunistas como bactérias, fungos, vírus e parasitas. A natureza oferece um banquete diversificado de desafios para as plantas, e cada espécie, em seu ecossistema particular, precisou encontrar soluções químicas únicas para lidar com seu ambiente [39].

Assim, ao longo da evolução, diferentes rotas biossintéticas foram desenvolvidas, permitindo que cada planta produzisse uma mistura personalizada de metabólitos secundários, sob medida para suas necessidades específicas. É como se cada planta fosse um laboratório químico, capaz de sintetizar moléculas complexas com propriedades farmacológicas e toxicológicas. Algumas destas substâncias podem ser verdadeiros venenos para herbívoros, enquanto outras podem ser utilizadas para atrair polinizadores ou para enviar sinais de alerta para plantas vizinhas em caso de perigo. De modo que, através dessas substâncias, as plantas se defendem, se comunicam e se adaptam, garantindo sua sobrevivência e perpetuação em um contexto natural de constantes mudanças [40].

A produção de metabólitos secundários representa uma estratégia de defesa e comunicação para as plantas. Dada sua natureza sésil, as plantas não podem recorrer à fuga física diante de ameaças como herbívoros, patógenos e plantas competidoras. A composição desses metabólitos não é estática, apresentando variações intra e interpoblacionais, bem como entre diferentes órgãos e estágios de desenvolvimento da planta. Essa diversidade molecular complexa é considerada uma estratégia evolutiva para evitar a seleção de herbívoros e patógenos especializados, que teriam maior dificuldade em se adaptar à complexidade química da planta [40].

A combinação de diferentes metabólitos secundários pode gerar um efeito sinérgico, potencializando suas atividades biológicas. Essa mistura complexa de compostos, como polifenóis e terpenoides, confere à planta uma proteção mais robusta e abrangente contra as diversas ameaças do ambiente. A produção de metabólitos secundários, portanto, representa uma adaptação evolutiva crucial para a

sobrevivência e o sucesso reprodutivo das plantas em seus respectivos ecossistemas. A complexidade e a dinamicidade da composição química dessas substâncias refletem a intrincada interação entre as plantas e seu ambiente, bem como a constante pressão seletiva exercida por herbívoros e patógenos [40, 41]. Nesse sentido, pode-se relacionar que a estrutura única de cada alcaloide é resultado direto das vias biossintéticas específicas que a planta utiliza para produzi-lo. Cada gênero e espécie vegetal possui um conjunto particular de enzimas e rotas metabólicas que moldam a estrutura final do alcaloide [41].

Essa diversidade biossintética se reflete na variedade de alcaloides encontrados em diferentes grupos de plantas. Por exemplo, o gênero *Sophora* é conhecido por produzir alcaloides quinolizidínicos e piperidínicos, enquanto o gênero *Erythrina* se destaca pela produção de alcaloides isoquinolínicos [10,11]. De modo que é possível observar diferenças nítidas nas sínteses das espécies dos gêneros antes citados, o que fica notório através dos aminoácidos distintos já presentes nas etapas iniciais das biossínteses apresentadas nas Figuras 1 e 2 [10, 11]. Essa especificidade na produção de alcaloides é uma característica distintiva de cada grupo de plantas, refletindo sua história evolutiva e suas adaptações únicas ao ambiente em que vivem.

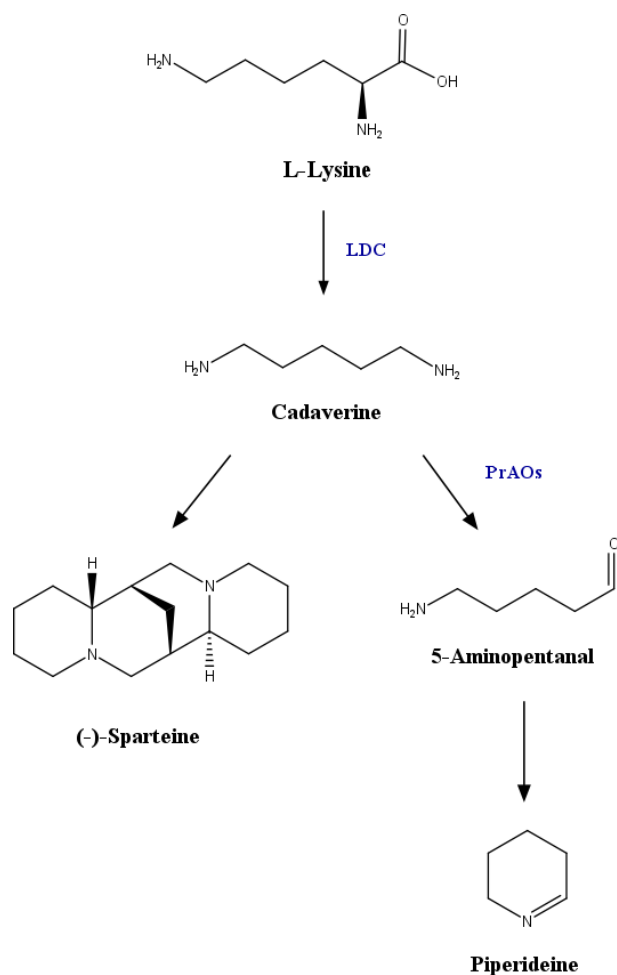


Figura 1. Biossíntese de alcaloides quinolizidínicos e piperidínicos de *Sophora* (adaptado de Huang et al. [10]).

A biossíntese dos alcaloides de *Sophora*, como a (-)-esparteína [(-)-sparteine], ocorre a partir da L-lisina, que é convertida em cadaverina (cadaverine) pela ação da lisina descarboxilase (LDC). A cadaverina segue duas rotas metabólicas: pode ser transformada diretamente em (-)-esparteína, um alcaloide quinolizidínico com propriedades farmacológicas e ecológicas, ou sofrer oxidação por poliamina oxidases (PrAOs), resultando em 5-aminopentanal, que se converte em piperideína (piperideine), um intermediário para a biossíntese de alcaloides piperidínicos [10].

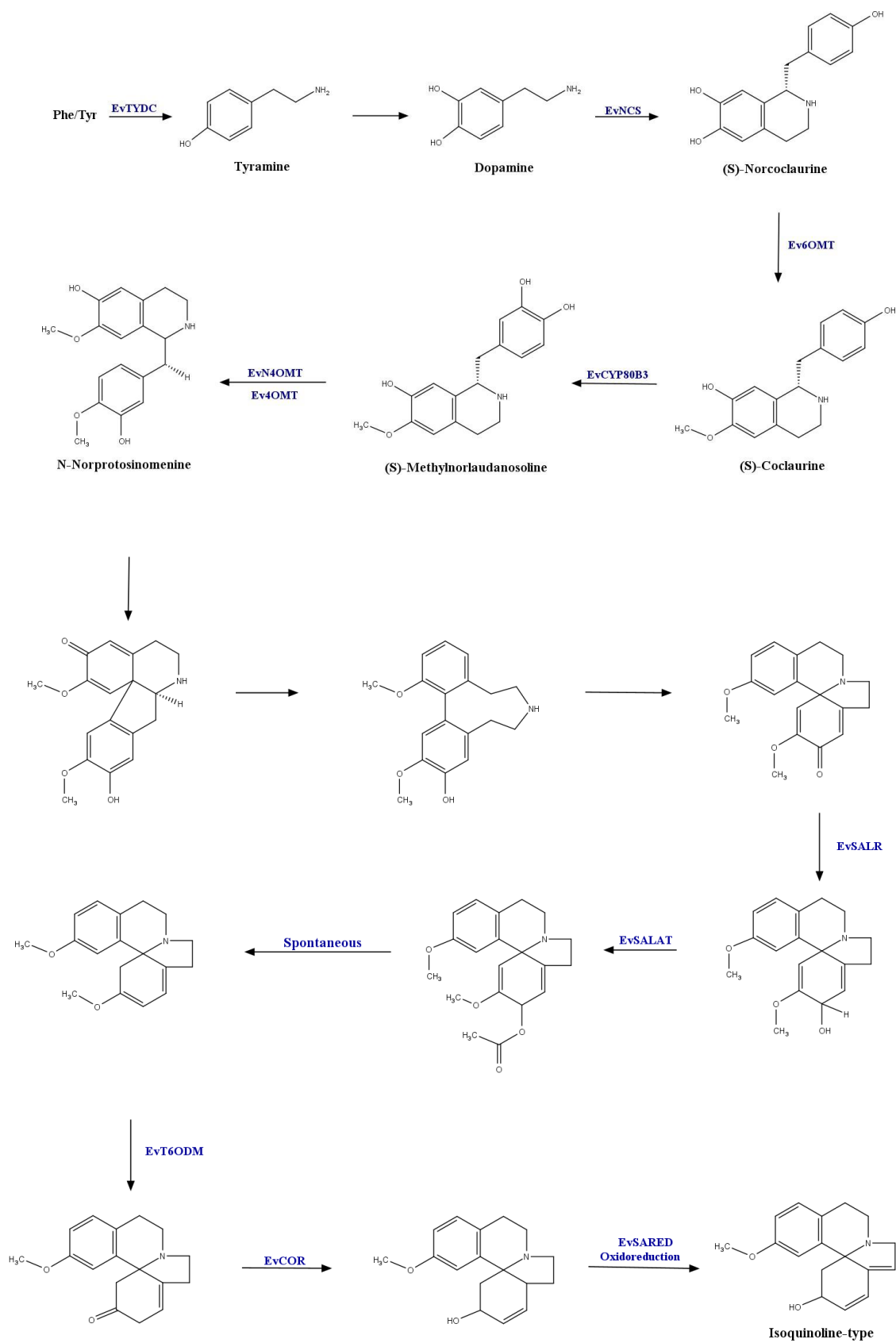


Figura 2. Biossíntese de alcaloides isoquinolínicos de *Erythrina* (adaptado de Chacon *et al* [11]).

A biossíntese dos alcaloides de *Erythrina* começa com a conversão da L-fenilalanina (Phe) ou L-tirosina (Tyr) em dopamina (dopamine), um precursor essencial na formação dos alcaloides isoquinolínicos. A dopamina reage com unidades de acetato para gerar (S)-norcoclaurina [(S)-norcoclaurine], que sofre modificações enzimáticas, como metilação e ciclização, resultando em intermediários como N-metilcoclaurina e cocalterina. A partir da cocalterina, ocorre a formação do núcleo tetracíclico característico desses alcaloides, um passo crucial catalisado por enzimas como a enzima ponte da berberina. As etapas finais incluem hidroxilações, metilações e oxidações, que diversificam os alcaloides de *Erythrina* [11].

Como antes mencionado, a biossíntese desses compostos não deve ser vista apenas como um mecanismo bioquímico isolado, mas como parte de uma dinâmica evolutiva, onde fatores genéticos e ambientais interagem para moldar a diversidade química observada na família Fabaceae [30, 105]. A distribuição desses metabólitos secundários reflete não apenas a proximidade genética entre os táxons, mas também eventos históricos de adaptação a nichos ecológicos específicos. Tornando assim a estrutura dos metabólitos uma útil ferramenta para relacionar a história evolutiva da família, com as moléculas produzidas por cada espécie [105, 106].

2.3 Quimiotaxonomia e mapas auto-organizáveis

Nesse mesmo sentido, a quimiotaxonomia, área que utiliza a análise de constituintes químicos para fins de classificação biológica, tem se consolidado como uma ferramenta valiosa na taxonomia vegetal. Ao empregar a identificação e análise de compostos químicos presentes em organismos, como metabólitos secundários, alcaloides, terpenoides e polifenóis, a quimiotaxonomia permite estabelecer relações taxonômicas e filogenéticas entre diferentes grupos de plantas. Essa abordagem, em sinergia com a análise de produtos naturais, oferece hipóteses aprofundadas sobre as relações evolutivas e adaptações ecológicas de espécies vegetais. A identificação de padrões químicos específicos, exclusivos de determinados grupos taxonômicos, possibilita a delimitação de espécies, gêneros e famílias, contribuindo para a construção de sistemas de classificação mais precisos e robustos [42].

A quimiotaxonomia, portanto, transcende a mera identificação de compostos químicos, representando uma abordagem multidisciplinar que integra dados químicos, morfológicos, genéticos e ecológicos. Ao considerar a complexidade da composição química das plantas, a quimiotaxonomia oferece uma perspectiva mais completa e abrangente sobre a diversidade e evolução do Reino Vegetal. No contexto da família Fabaceae (Leguminosae), notória por sua rica diversidade química e importância ecológica, a quimiotaxonomia tem se mostrado particularmente útil [42].

A análise de alcaloides, flavonoides e outros compostos presentes nessas plantas tem auxiliado na elucidação de relações taxonômicas complexas, contribuindo para a revisão e aprimoramento da classificação dessa importante família de angiospermas. Em suma, a quimiotaxonomia representa uma ferramenta poderosa na taxonomia vegetal, oferecendo uma abordagem multidimensional que integra dados químicos, morfológicos, genéticos e ecológicos. Ao desvendar a complexidade da composição química das plantas, a quimiotaxonomia contribui para a construção de sistemas de classificação mais precisos e elucidativos, aprofundando nossa compreensão sobre a diversidade e a evolução do mundo vegetal [42].

No cenário da quimiotaxonomia, os Mapas Auto-organizáveis (do inglês, SOMs) emergem como uma ferramenta analítica de grande relevância. Sua capacidade de processar e interpretar perfis químicos complexos, especialmente aqueles derivados de metabólitos secundários como alcaloides, possibilita a geração de representações visuais intuitivas e informativas. Esses mapas revelam padrões intrincados e relações sutis entre amostras, permitindo a identificação de espécies com similaridades químicas notáveis, mesmo que estas pertençam a grupos taxonômicos distintos e distantes na árvore filogenética. Essa característica singular contribui de forma significativa para a construção de classificações mais precisas e robustas, aprimorando nossa compreensão das relações filogenéticas entre as plantas [43].

Do ponto de vista filogenético, os gêneros *Erythrina* e *Sophora* ocupam clados bem definidos dentro da família Fabaceae, refletindo uma longa trajetória evolutiva marcada por processos de especiação e adaptação. Estudos baseados em filogenia molecular, como o conduzido por Zhao *et al.* [6], proporcionaram avanços significativos na compreensão das relações evolutivas entre esses gêneros e outros membros da família, permitindo uma reconstrução mais detalhada dos eventos que moldaram sua diversificação [6, 18].

A determinação precisa do posicionamento filogenético de *Erythrina* e *Sophora* é fundamental para elucidar a evolução de seus traços morfológicos, fisiológicos e bioquímicos, pois possibilita a identificação de características herdadas de ancestrais comuns e das inovações que surgiram ao longo da evolução. Além disso, essa abordagem permite inferências mais robustas sobre a origem, biossíntese e diversificação dos alcaloides encontrados nessas plantas, que desempenham funções ecológicas e farmacológicas relevantes [17]. Dessa forma, a filogenia molecular não apenas esclarece a história evolutiva desses gêneros, mas também oferece um arcabouço essencial para estudos em quimiosistemática, bioprospecção e conservação da biodiversidade. A Figura 3 demonstra a posição dos gêneros *Erythrina* e *Sophora* na filogenia da família Fabaceae, evidenciando a distância entre os clados de cada um dos gêneros, e por consequência inferindo possíveis divergências evolutivas importantes [6, 19].

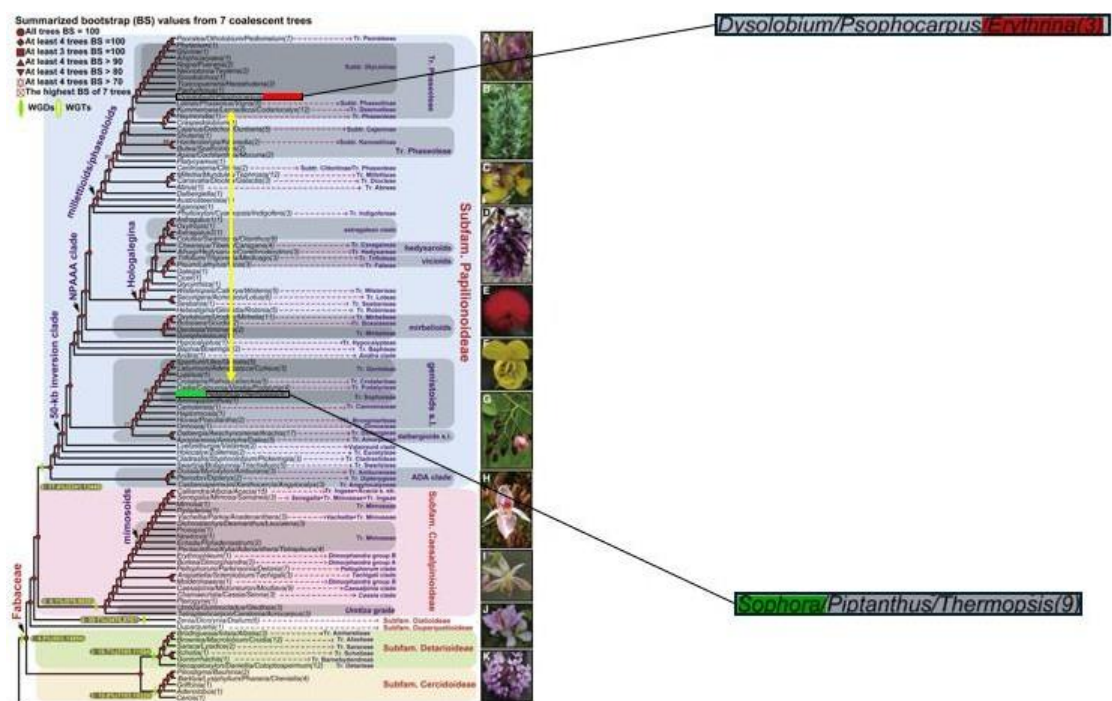


Figura 3. Uma visão geral da filogenia de Fabaceae (adaptado de Zhao *et al.* [6])

Ademais, os SOMs oferecem a oportunidade de identificar marcadores químicos específicos, que são característicos de determinados grupos taxonômicos. Essa aplicação facilita a descoberta de novos táxons, além de aprofundar o

conhecimento sobre a evolução química das plantas ao longo do tempo e sua adaptação a diferentes ambientes e pressões seletivas. Em outras palavras, os SOMs permitem aos pesquisadores visualizar e analisar grandes conjuntos de dados químicos de forma mais eficiente, revelando padrões que seriam difíceis de identificar por outros métodos. Essa capacidade de "visualização" da complexidade química das plantas é fundamental para avançar na pesquisa em quimiotaxonomia e áreas relacionadas. Em resumo, os SOMs representam uma ferramenta poderosa na quimiotaxonomia, auxiliando na análise de dados complexos e na identificação de padrões químicos significativos. Sua utilização contribui para aprimorar a classificação de plantas, desvendar relações evolutivas e aprofundar nossa compreensão sobre a diversidade química do Reino Vegetal [43].

2.4 Redes neurais e mapas auto-organizáveis

As redes neurais artificiais (RNAs) são sistemas computacionais projetados para simular o comportamento dos neurônios biológicos presentes no cérebro humano. Inspiradas na neurociência, as RNAs são compostas por unidades interconectadas denominadas neurônios artificiais [82]. O funcionamento das RNAs é baseado no processo de aprendizado supervisionado. Durante o treinamento, a rede é exposta a um conjunto de dados rotulados, onde as entradas são associadas a saídas conhecidas. O objetivo do treinamento é minimizar o erro entre as saídas previstas pela rede e as saídas reais, ajustando os pesos das conexões entre os neurônios. Esse ajuste é realizado utilizando algoritmos de otimização, como o gradiente descendente. Uma vez treinada, a rede neural é capaz de generalizar seu conhecimento para novas entradas, realizando previsões e classificações com base no aprendizado anterior [82, 83].

Os mapas auto-organizáveis, também conhecidos como mapas de Kohonen, são um tipo de rede neural artificial de aprendizado não supervisionado. Os SOMs se distinguem pela sua capacidade de organizar dados de alta dimensão em um espaço de baixa dimensão de forma não supervisionada [77]. Esta característica torna-os uma ferramenta valiosa para tarefas de visualização e clustering de dados, através do treinamento, os SOMs ajustam suas unidades (neurônios) de modo a refletir a topologia dos dados de entrada, preservando a estrutura e as relações intrínsecas entre os dados [77].

Os neurônios em mapas auto-organizáveis são unidades fundamentais que compõem a grade onde os dados são mapeados. Cada neurônio é representado por um vetor de peso, que é ajustado durante o processo de treinamento para refletir as características dos dados de entrada. O conceito de neurônio em SOMs está intimamente ligado à ideia de aprendizado competitivo e organização topológica [77, 78].

Durante o treinamento de um SOM, cada ponto de dados de entrada é comparado a todos os neurônios da grade usando uma métrica de distância, como a distância euclidiana [78]. O neurônio cujo vetor de peso está mais próximo do ponto de dados de entrada é chamado de "neurônio vencedor" (do inglês, BMU). Após identificar o BMU, os pesos desse neurônio e dos neurônios vizinhos são ajustados para se aproximarem ainda mais do ponto de dados de entrada [78].

Esse processo de ajuste iterativo permite que os SOMs capturem a estrutura topológica dos dados, garantindo que pontos de dados semelhantes sejam mapeados para neurônios próximos na grade. A organização dos neurônios em uma grade bidimensional facilita a visualização de padrões e agrupamentos nos dados de alta dimensão. A inspiração neurofisiológica dos SOMs vem da observação de que diferentes entradas sensoriais no cérebro são mapeadas em áreas correspondentes no córtex cerebral. Esse mapeamento topográfico é semelhante ao modo como os SOMs organizam os dados de entrada em uma grade de neurônios [77, 79].

Para melhor entendimento, pode-se aplicar a teoria que fundamenta os SOMs a uma situação real do cotidiano, como por exemplo, um supermercado que necessita organizar os dados dos seus clientes e relacioná-los ao comportamento de compra de cada um deles. Serão considerados quatro dados para realizar a construção desse mapa: a renda, a idade, a frequência de compra, e o valor gasto a cada visita ao estabelecimento. Inicialmente, é necessário normalizar os valores dos dados disponíveis, afinal é notório que a renda (geralmente em milhares ou milhões de reais) e a idade (geralmente em dezenas de anos) são valores que não possuem a mesma escala numérica [79, 105].

Na normalização, a ideia é transportar os valores dos dados para uma escala numérica semelhante, evitando assim que a magnitude de algum deles domine a análise de forma isolada. Logo após o processo de normalização, é criado um mapa bidimensional com células (espaços unitários) que são chamadas de neurônios, a cada um desses espaços unitários é atribuído um vetor peso de forma aleatória, esse

vetor peso servirá para comparação com os dados iniciais normalizados, de modo que cada vetor possui a mesma quantidade elementos que a quantidade de dados iniciais, para situação hipotética atual cada vetor peso tem quatro elementos [105].

Após a atribuição de todos os pesos para cada um dos neurônios presentes, é iniciado o processo de comparação dos vetores de entrada (dados dos clientes) com os vetores gerados para cada neurônio do mapa. O processo de comparação é geralmente realizado através da distância euclidiana entre os vetores de entrada e os vetores peso, de modo que o vetor de menor distância euclidiana concede ao neurônio a qual está localizado o título de neurônio vencedor [78, 105]. A distância euclidiana pode ser calculada segundo a fórmula abaixo.

$$D = \sqrt{\sum_{i=1}^n (x_i - w_i)^2} \quad (1)$$

* D = distância euclidiana

* n = dimensão dos vetores

* x_i = valor da i -ésima característica no vetor de entrada

* w_i = valor da i -ésima característica no vetor peso dos neurônios

Na próxima etapa, o vetor peso do BMU e dos neurônios vizinhos são calibrados para se aproximar o máximo possível dos dados de entrada, essa calibração ocorre em maior magnitude para o BMU e sua intensidade vai diminuindo ao passo que é feita para neurônios mais distantes do vencedor, ou seja, neurônios próximos ao BMU serão calibrados em magnitudes semelhantes. O passo anterior é repetido uma série de vezes, passando por diferentes conjuntos dos dados de entrada, com o objetivo de organizar os neurônios e estruturá-los em um padrão semelhante ao dos dados de entrada originais [78, 79].

Por fim, é gerado um mapa bidimensional onde neurônios próximos representam dados com padrões semelhantes, essa proximidade é baseada nos dados de comparação construídos através da distância euclidiana. Sendo assim, clientes com comportamento de compra semelhante ocupam neurônios próximos no mapa, por consequência, clientes com comportamento de compra muito diferentes ocupam neurônios distantes. Possibilitando assim uma organização visual e geral dos

clientes através de um mapa bidimensional, facilitando a observação de padrões através da redução da dimensionalidade dos dados analisados [79].

Como os SOMs são utilizados para organizar e visualizar dados de alta dimensão em um espaço de baixa dimensão, eles são particularmente úteis na análise de dados químicos, de modo que esses podem ser usados para agrupar compostos com base em suas propriedades e estruturas químicas. Ao mapear compostos similares próximos uns dos outros, os SOMs facilitam a identificação de padrões e relações entre diferentes moléculas [80]. Nesse contexto, pode-se entender que os SOMs podem ser utilizados para agrupar metabólitos secundários de acordo com suas propriedades químicas e estruturais, de modo a permitir que compostos com estruturas químicas semelhantes sejam mapeados para neurônios próximos na grade. Isso facilita a visualização de padrões e a identificação de grupos de metabólitos com propriedades similares [22, 80].

A aplicação de SOMs na organização de metabólitos secundários é uma abordagem promissora para a química de produtos naturais, de modo que ao facilitar a análise de grandes conjuntos de dados químicos, os SOMs contribuem para o avanço do conhecimento sobre a diversidade química e as funções biológicas dos metabólitos secundários [81].

2.5 Validação de modelos de aprendizado de máquina e análise de componente principal

O processo de validação de modelos de aprendizado de máquina é um componente crítico no desenvolvimento de algoritmos robustos e precisos. A correta validação permite garantir que o modelo treinado não apenas se ajusta bem aos dados de treino, mas também se generaliza efetivamente para novos dados. A partição do banco de dados em treino e teste, se apresenta como um método bastante utilizado para validação de modelos de aprendizado de máquina, sua aplicação consiste dividir o conjunto de dados disponível em dois subconjuntos, chamados de treino e teste, essa técnica é conhecida como validação de Holdout [48, 51]. A Figura 4 apresenta uma visão geral para formulação de modelos de aprendizado de máquina.

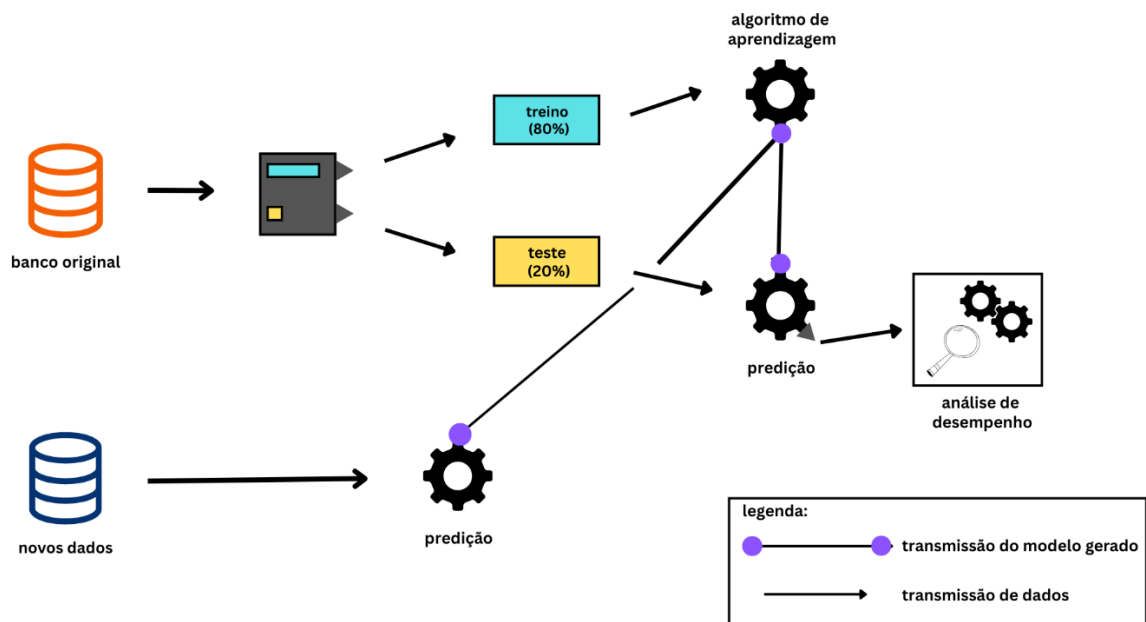


Figura 4. Visão geral da formulação de modelos de aprendizado de máquina (adaptado de Menezes *et al.* [51]).

O modelo aprende através do subconjunto treino, de modo que, durante esta fase o algoritmo ajusta seus parâmetros internos para minimizar determinados erros nos dados de treino. Já através do subconjunto de teste o modelo é avaliado, afinal este conjunto de dados não foi visto pelo modelo durante a fase de treinamento e, portanto, fornece uma estimativa imparcial da performance, capacidade preditiva, e qualidade. Porém é necessário salientar que essa técnica, por se utilizar de apenas uma única divisão dos dados, não é suficiente para validação [48, 51].

Nesse sentido, pode-se destacar a validação cruzada (do inglês, *cross-validation*), que se utiliza da técnica de *k* partes (do inglês, *k-fold*), de modo a dividir o banco em diversos subconjuntos de treino e teste, realizando assim validações de acordo com a quantidade *k*, ou seja no 5-fold, tem-se a divisão de cinco grupos de treino e teste que serão utilizados para validar o modelo, de modo que os subconjuntos são escolhidos de forma aleatória promovendo a formação de grupos de treino e teste distintos para cada uma das divisões realizadas [44, 45].

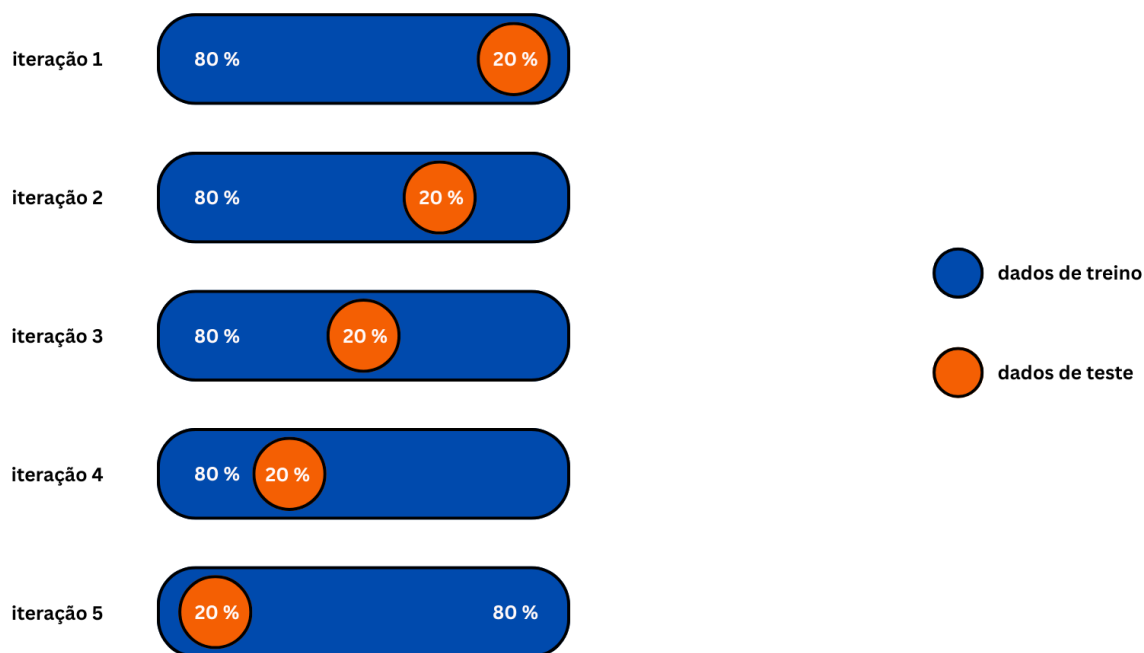


Figura 5. Ilustração da divisão de dados (subconjuntos de treino e teste) realizada para a validação cruzada do tipo 5-fold.

Contudo, é necessário entender que existem duas formas de validação principais nesse contexto, que se diferem em relação a origem dos dados utilizados para o teste, essas são a validação interna e a externa. Na validação interna, avalia-se o modelo utilizando o mesmo conjunto de moléculas que foi usado para treiná-lo, chamado de conjunto de treino. Já na validação externa, verificamos o desempenho do modelo com um novo conjunto de moléculas, que ele ainda não tinha visto durante o treinamento. Portanto, a validação cruzada é uma técnica essencial para medir a capacidade de um modelo de se generalizar para novos dados. Nesse método, o modelo é testado várias vezes, e ao final, o desempenho é determinado pela média aritmética das avaliações realizadas [46, 47].

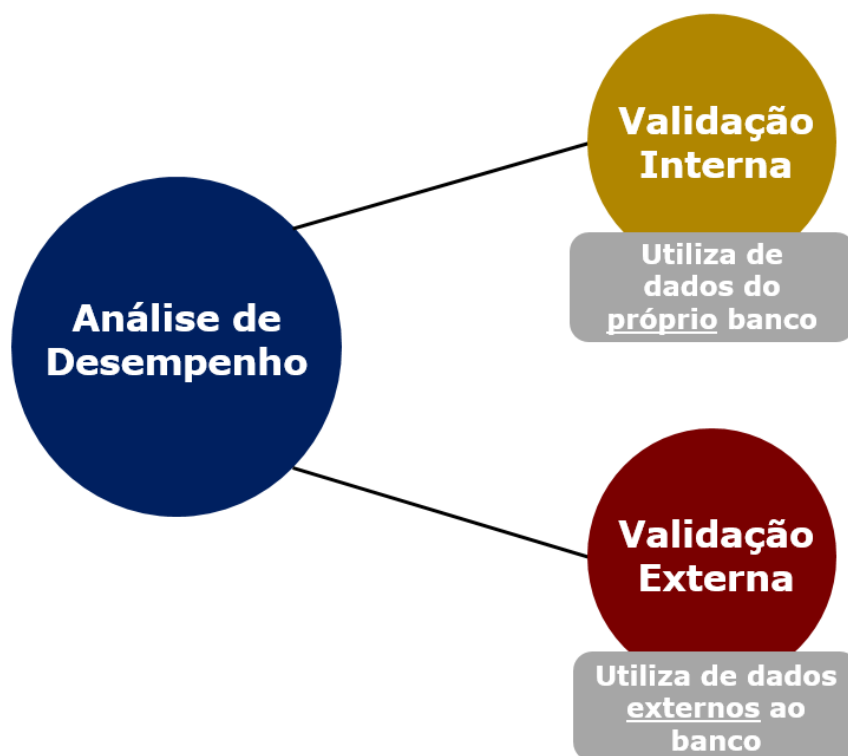


Figura 6. Comparação das validações interna e externa (adaptado de Menezes et al. [51])

Para a validação cruzada interna de k partes, o conjunto de treino é dividido aleatoriamente e sem reposição em k partes iguais. O processo de validação consiste em realizar k repetições, em cada iteração, $k-1$ partes são usadas para treino e criação do modelo, enquanto a parte restante é utilizada para teste. Em cada nova repetição, uma parte diferente é usada para teste, enquanto as outras $k-1$ são usadas para treino. Dessa forma, ao final do processo, todas as k partes terão sido usadas como conjunto de teste, permitindo uma avaliação completa do modelo. A performance final do modelo é obtida pela média dos resultados das k iterações realizadas. Já para a validação externa, a escolha aleatória dos subconjuntos geralmente não se apresenta como a melhor opção, nesse caso recomenda-se o método da validação cruzada estratificada. Nesta abordagem, o modelo é desenvolvido já integrando a validação cruzada, o conjunto de treino é dividido em k partes de forma estratificada, e cada modelo é avaliado por um conjunto de teste composto por moléculas desconhecidas para o modelo, caracterizando, assim, a validação externa. Ao final do processo, tem-se modelos que foram validados separadamente, de modo que a performance final é obtida através da média aritmética de todos os modelos formados. Geralmente na

validação cruzada de k , é utilizada a divisão em cinco ou dez partes, promovendo assim um equilíbrio satisfatório entre custo computacional e precisão do método de validação do modelo [49, 50].

A Análise de Componentes Principais (ACP) é uma técnica estatística amplamente utilizada para reduzir a dimensionalidade de conjuntos de dados. Essencialmente, ela transforma um grande conjunto de variáveis possivelmente correlacionadas em um conjunto menor de variáveis não correlacionadas, chamadas de componentes principais. Esta transformação é realizada de tal maneira que o primeiro componente principal possui a maior variação possível nos dados, e cada componente subsequente possui a maior variação possível, sob a restrição de ser ortogonal aos componentes anteriores [52].

A formação dos componentes principais é baseada na decomposição dos dados através da matriz de covariância ou da matriz de correlação. Inicialmente, os dados são padronizados para garantir que cada variável tenha igual influência na análise. Em seguida, a matriz de covariância é calculada e, através da decomposição em valores singulares (SVD) ou da decomposição espectral, são extraídos os autovalores e os autovetores. Os autovalores representam a magnitude da variação explicada por cada componente, enquanto os autovetores representam a direção dos componentes principais nos dados originais [52].

A relação entre QSAR (do inglês, Quantitative Structure-Activity Relationship) e ACP é bastante significativa, especialmente no campo da química medicinal e da bioinformática. O QSAR é uma metodologia que busca correlacionar a estrutura química de compostos com suas atividades biológicas ou propriedades físico-químicas. Como supracitado, a ACP, por sua vez, é uma técnica estatística utilizada para reduzir a dimensionalidade dos dados, facilitando a visualização e interpretação dos mesmos [53].

Uma das principais aplicações do QSAR junto à ACP é a simplificação dos dados estruturais. Em estudos de QSAR, é comum lidar com um grande número de descritores moleculares que representam diferentes aspectos das estruturas químicas. A ACP pode ser utilizada para reduzir a dimensionalidade desses descritores, transformando-os em um conjunto menor de componentes principais que retêm a maior parte da variação dos dados originais. Isso facilita a construção de modelos QSAR mais robustos e interpretáveis [51].

Além disso, a ACP pode ser aplicada na seleção de variáveis em modelos QSAR. Ao identificar os componentes principais que mais contribuem para a variação nos dados, é possível selecionar os descritores moleculares mais relevantes para a atividade biológica em estudo. Isso não apenas melhora a precisão dos modelos QSAR, mas também reduz o tempo de processamento e os recursos computacionais necessários para a análise [51 ,53].

2.6 Doenças Tropicais Negligenciadas: Leishmaniose

As doenças tropicais negligenciadas (NTDs) englobam um grupo de 20 condições que são predominantes em regiões tropicais e afetam principalmente populações carentes, com um impacto desproporcional em mulheres e crianças [60]. A parcela da população mundial que pode ser afetada por essas doenças continua alta, resultando em consequências sociais e econômicas para cerca de 20% da população global, o que mostra que as NTDs ainda são um problema significativo de saúde pública [61, 62]. Um dos principais desafios para controlar e eliminar essas doenças é a falta de interesse da maioria das empresas farmacêuticas, uma vez que a pesquisa e produção de novos medicamentos para tratá-las não gerariam lucros significativos. Os tratamentos disponíveis atualmente possuem várias limitações, como baixa eficácia, alta toxicidade e durações prolongadas [60, 61].

Dentro desse contexto, pode-se destacar a leishmaniose, uma doença endêmica em várias regiões tropicais e subtropicais do mundo e que se apresenta em duas formas principais: leishmaniose cutânea e leishmaniose visceral. Enquanto a primeira causa lesões na pele, a segunda, conhecida como calazar, afeta órgãos internos como fígado e baço, podendo ser fatal se não tratada adequadamente [64, 65]. A leishmaniose visceral, em particular, é uma das formas mais graves, demandando maior atenção quanto ao desenvolvimento de novos tratamentos. Atualmente, os tratamentos disponíveis para a leishmaniose são limitados e frequentemente ineficazes. Os medicamentos mais comuns, como os antimoniais pentavalentes, apresentam várias limitações, incluindo alta toxicidade, efeitos colaterais severos e administrações dolorosas [66]. Além disso, a resistência aos medicamentos vem aumentando, tornando muitos dos tratamentos tradicionais cada vez menos eficazes. Este cenário revela a urgente necessidade de desenvolvimento

de novos fármacos que sejam mais seguros, eficientes e acessíveis para a população afetada [66, 67].

A pesquisa e o desenvolvimento de novos fármacos para a leishmaniose enfrentam diversos desafios, entre eles, destaca-se o reduzido interesse da indústria farmacêutica em investir em doenças que predominam em regiões economicamente desfavorecidas [60, 61]. A pesquisa de novos medicamentos exige investimentos significativos e, muitas vezes, não oferece retorno financeiro compatível, o que leva à marginalização das doenças tropicais negligenciadas, como a leishmaniose [60, 61, 66]. A identificação de novos alvos moleculares para o desenvolvimento de fármacos, bem como a identificação de compostos naturais com atividade leishmanicida, tem sido o foco de diversas pesquisas desenvolvidas ao redor do mundo, a criação de bancos de dados de compostos promissores e o compartilhamento de informações entre pesquisadores são cruciais para acelerar o desenvolvimento de novas terapias [67].

2.7 Farmacocinética e Propriedades ADMET

A farmacocinética é uma área fundamental da farmacologia que estuda o caminho que os medicamentos percorrem no organismo, desde a administração até a eliminação. Esse estudo é dividido em quatro etapas principais: absorção, distribuição, metabolismo e excreção (ADME). A absorção refere-se ao processo pelo qual o medicamento é transferido do local de administração para a corrente sanguínea. A distribuição envolve o transporte do medicamento pelo corpo, onde pode se ligar a proteínas plasmáticas ou se acumular em tecidos específicos. O metabolismo, principalmente no fígado, transforma o medicamento em metabólitos mais fáceis de excretar. Finalmente, a excreção é o processo pelo qual os resíduos do medicamento são eliminados do corpo, geralmente pelos rins [54, 55].

As propriedades ADMET (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade) são cruciais no desenvolvimento de novos fármacos. Essas propriedades determinam a eficácia e a segurança de um medicamento, influenciando sua biodisponibilidade e potencial de causar efeitos adversos. A avaliação dessas propriedades é essencial para garantir que um novo medicamento atinja seu alvo terapêutico de maneira eficaz e segura. A toxicidade, em particular, é uma

consideração crítica, pois um medicamento eficaz deve ter um perfil de segurança aceitável para ser viável clinicamente [54, 56].

No design de fármacos, a compreensão das propriedades ADMET é vital para a otimização de novos compostos. A avaliação precoce dessas propriedades pode reduzir significativamente a taxa de falha no desenvolvimento de medicamentos, economizando tempo e recursos. Por exemplo, um composto com baixa absorção ou rápida eliminação pode não alcançar concentrações terapêuticas adequadas no organismo.

Da mesma forma, um medicamento que se distribui amplamente em tecidos não-alvo pode causar efeitos colaterais indesejados [54, 57, 58]. É importante salientar também que as plataformas de ADMET utilizam a modelagem molecular *in silico* ferramenta para prever as propriedades farmacocinéticas e toxicológicas de novos compostos antes mesmo de serem testados experimentalmente. A modelagem *in silico* envolve o uso de algoritmos computacionais e simulações para analisar a interação de moléculas com alvos biológicos, e assim prever as propriedades farmacocinéticas dos compostos. Essas técnicas permitem a triagem rápida e eficiente de grandes bibliotecas de compostos, identificando aqueles com maior potencial terapêutico e menores riscos de efeitos adversos [58, 59].

2.8 Docking Molecular

A inibição de um alvo terapêutico envolve um esquema de “chave e fechadura”, onde os resíduos dessa macromolécula biológica se ligam com porções estruturais de um ligante através de interações dos mais diversos tipos (ligações de hidrogênio, π - π , doador-doador, hidrofóbicas etc). Ao propor um lead, ou seja, uma molécula potencialmente ativa contra um alvo terapêutico, é imprescindível conhecer a extensão e a natureza dessas interações [84]. Nessa perspectiva, o docking molecular se enquadra no grupo de métodos computacionais baseados na estrutura, que se caracterizam pelo baixo custo de operação e alta confiabilidade [85].

O docking simula um processo conhecido como reconhecimento molecular, ou seja, a formação de um complexo entre uma macromolécula biológica (proteínas) e ligantes, ou seja, moléculas pequenas que interagem com o alvo por ligações não-covalentes. Uma vez formado o complexo com o ligante existe um outro aspecto a ser considerado, a especificidade. Para ocorrer a inibição, o ligante precisa interagir na

posição certa e com as interações certas, o que vai depender da afinidade específica alvo-ligante [86].

No processo de reconhecimento molecular, diversos fatores como interações entálpicas e ambiente proteico, influenciam na qualidade da interação ligante-alvo. Portanto, algumas simplificações são feitas para contornar essa complexidade, mas preservando a afinidade química e a especificidade. Vários algoritmos são utilizados no docking para prever as possíveis conformações do complexo binário, pontuando cada uma dessas conformações dos ligantes através de um somatório de todas as interações obtidas, fornecendo a magnitude quantitativa da interação ligante-alvo [86].

Uma vez selecionados o(s) ligante(s), o alvo é preparado, a partir da retirada de moléculas de água e cofatores iônicos ou não iônicos que possam estar previamente ligados. Uma vez preparadas, as moléculas são submetidas ao docking, seguindo um conjunto de parâmetros altamente configuráveis, a critério do analista. As pontuações obtidas são analisadas e as conformações (ou poses) com menor energia, ou seja, com melhor estabilidade nas interações, são selecionadas como as melhores candidatas. Essas moléculas selecionadas podem então seguir para demais investigações como testes *in vitro*, o que faz o docking ser um método bastante atrativo por diminuir radicalmente os custos financeiros e de tempo para o design de novos fármacos [87].

Nas últimas décadas, o desenvolvimento de novos softwares para a aplicação do docking atingiu ritmo acelerado. Os mais usados são o AutoDock, Dock, FlexX, Glide, Gold, Suflex, ICM, LigandFit, entre outros [87]. Algoritmos como o MolDock Scoring Function e PLANTS geram a energia total (Escore) das interações do complexo a partir da soma entre a energia da interação ligante-alvo (Einter) e a energia interna do ligante (Eintra).

Pela natureza aditiva de geração das funções de pontuação, o peso molecular se mostra um fator que influencia dramaticamente na classificação das posturas, resultando em uma tendência de designar estruturas moleculares mais pesadas para topo da lista classificatória. Para contornar esse problema, métodos de normalização dos scores podem ser aplicados em detrimento dos valores de peso molecular, assim como o ponderamento que conceda diferentes pesos para a função de score aplicada [88].

Nesse sentido, e entendendo a relevância do docking para o desenvolvimento de fármacos leishmanicidas, a ornitina descarboxilase (ODC) pode ser destacada, pois é uma enzima chave na via de biossíntese de poliaminas, catalisando a conversão da ornitina em putrescina, um intermediário essencial na produção de espermidina e espermina, moléculas cruciais para processos biológicos como crescimento celular, replicação do DNA e estabilização de estruturas macromoleculares. Nas espécies do gênero *Leishmania*, a dependência dessas vias metabólicas é particularmente acentuada, uma vez que os parasitas utilizam as poliaminas para manter sua viabilidade e capacidade de proliferação no ambiente hostil do interior dos macrófagos do hospedeiro vertebrado. A interrupção dessa rota metabólica, portanto, representa uma estratégia eficaz para comprometer a sobrevivência do parasita, tornando a ODC um alvo bioquímico de grande interesse no desenvolvimento de terapias antileishmaniose [71, 109].

A relevância da ODC como alvo farmacológico se fortalece ainda mais pelo fato de que a enzima da *Leishmania* apresenta diferenças significativas em sua estrutura tridimensional e nos mecanismos regulatórios em comparação com a ODC humana. Essa divergência estrutural permite que inibidores seletivos sejam projetados para se ligarem especificamente à enzima do parasita, reduzindo o risco de efeitos tóxicos ao hospedeiro. Essa seletividade é uma vantagem crítica, já que os tratamentos atuais contra a leishmaniose frequentemente apresentam alta toxicidade e eficácia limitada. Com isso, a ODC surge como uma alternativa viável e promissora para a descoberta de novos fármacos mais seguros e eficazes, particularmente quando associada a abordagens computacionais como o docking molecular, que permitem explorar racionalmente o desenho de moléculas inibidoras com alta afinidade e especificidade [71, 109].

Outra enzima que é de importante análise no processo de desenvolvimento de fármacos para Leishmaniose é a 14-alfa-desmetilase (CYP51). A CYP51 desempenha um papel central na biossíntese do ergosterol, um esteroide essencial para a manutenção da membrana plasmática de protozoários como *Leishmania spp.*. O ergosterol garante a fluidez, estabilidade e funcionalidade da membrana celular, sendo vital para a sobrevivência do parasita. Nos mamíferos, esse papel é desempenhado pelo colesterol, o que oferece uma oportunidade de seletividade terapêutica: ao inibir a CYP51 da *Leishmania*, é possível interromper a formação de ergosterol sem interferir significativamente nas vias esteroidais do hospedeiro. Isso torna a enzima

um alvo estratégico para o desenvolvimento de novos agentes antileishmaniais. Além disso, o docking molecular explora as diferenças estruturais entre a CYP51 da *Leishmania* e a homóloga humana, possibilitando o desenvolvimento de inibidores seletivos com menor toxicidade [72, 110].

Nessa mesma perspectiva, pode-se destacar também a PnC1. A PnC1 (nicotinamidase ou piridina-nucleosídeo hidrolase 1) é uma enzima envolvida no metabolismo da nicotinamida (vitamina B3) e na regeneração de NAD^+ via a via de salvamento. Essa via é essencial para a manutenção do metabolismo energético celular e para a regulação do estresse oxidativo — processos críticos para a sobrevivência e virulência de *Leishmania spp.*. Diferente dos mamíferos, que não possuem nicotinamidase funcional, *Leishmania* depende da PnC1 para converter nicotinamida em ácido nicotínico, um precursor essencial para a biossíntese de NAD^+ . Essa diferença metabólica torna a PnC1 um alvo altamente promissor e seletivo para o desenvolvimento de novos fármacos antiparasitários [73, 111].

Nesse contexto, o docking molecular desempenha um papel-chave na exploração racional da PnC1 como alvo terapêutico. A técnica permite modelar a interação entre a enzima e potenciais inibidores, simulando como pequenas moléculas podem se ligar ao sítio ativo da PnC1 de *Leishmania*. Isso possibilita a identificação de compostos que bloqueiem a função enzimática, interrompendo a regeneração de NAD^+ e levando à morte celular do parasita. Além disso, o docking ajuda a analisar as diferenças estruturais entre a PnC1 parasitária e outras enzimas humanas relacionadas, favorecendo o desenvolvimento de inibidores seletivos com menor toxicidade [73, 111].

A piridoxal quinase (PLK) é uma enzima essencial no metabolismo da vitamina B6, catalisando a fosforilação de piridoxal, piridoxamina e piridoxina para formar piridoxal-5'-fosfato (PLP), a forma ativa da vitamina. O PLP atua como cofator em numerosas reações enzimáticas fundamentais, incluindo o metabolismo de aminoácidos e a biossíntese de neurotransmissores. Em *Leishmania spp.*, a PLK é indispensável para o crescimento e diferenciação do parasita, pois regula vias metabólicas vitais que não possuem rotas alternativas. Essa dependência da vitamina B6 ativa, somada às diferenças estruturais entre a PLK do parasita e a homóloga humana, torna essa enzima um alvo estratégico para o desenvolvimento de fármacos seletivos [74, 112].

Nesse cenário, o docking molecular também se apresenta como uma ferramenta importante para explorar a piridoxal quinase como alvo terapêutico. Essa técnica permite simular como moléculas candidatas se encaixam no sítio ativo da PLK de *Leishmania*, identificando ligantes com alta afinidade capazes de inibir sua atividade. O docking também permite avaliar seletividade, examinando como os mesmos compostos interagem com a PLK humana, o que é crucial para reduzir efeitos colaterais. Assim, o uso do docking molecular acelera a triagem e o desenho racional de inibidores da PLK, contribuindo para o avanço de terapias mais eficazes contra a leishmaniose [74, 112].

CAPÍTULO 3: OBJETIVOS

3. OBJETIVOS

3.1 Objetivo Geral

Estabelecer um sistema de classificação para alcaloides de dois gêneros da família Fabaceae, *Erythrina* e *Sophora*, assim como selecionar alcaloides que possuam potenciais propriedades para serem novos fármacos para leishmaniose.

3.2 Objetivos Específicos

- Construir um banco com os dados quimiotaxonômicos de alcaloides pertencentes a família Fabaceae;
- Gerar os descritores moleculares dessas moléculas, assim como utilizá-los para construção do sistema de classificação;
- Gerar as estruturas tridimensionais das moléculas, e utilizá-las no docking molecular;
- Predizer as propriedades farmacocinéticas dos compostos mais promissores obtidos através do docking.

CAPÍTULO 4: METODOLOGIA

4. METODOLOGIA

4.1 Banco de dados dos alcaloides

Com base em uma pesquisa bibliográfica abrangente, foi estabelecida uma coleção de moléculas de alcaloides isoladas da família Fabaceae. A pesquisa, conduzida eletronicamente entre 2019 e 2024, utilizou tanto o SciFinder (<https://scifinder.cas.org/>) quanto o Web of Science (<https://clarivate.com/products/web-of-science/>). Esse banco de dados compilado foi subsequentemente incorporado à ferramenta online SistematX [12]. Ele engloba as estruturas químicas, códigos SMILES, nomes dos compostos, juntamente com referências bibliográficas e informações taxonômicas detalhadas. Essas informações incluem a classificação da espécie (da família até a espécie individual) e a localização geográfica onde os compostos foram isolados.

4.2 Descritores Moleculares

As estruturas dos alcaloides foram inicialmente representadas como códigos SMILES e inseridas nos softwares MarvinSketch [69]. Posteriormente, o software Standardizer foi empregado para padronizar essas estruturas [91]. Esse processo envolveu a conversão das diversas estruturas químicas em uma representação canônica consistente, adicionando átomos de hidrogênio, aromatizando as estruturas, gerando representações 2D e, finalmente, salvando os compostos em formato SDF. Uma vez que os dados foram processados e organizados pelo Standardizer, as estruturas 2D foram utilizadas como entrada para o software alvaDesc 2.0 [3]. Logo após, foram calculados um total de 104 diferentes descritores para cada uma das 650 moléculas presentes no banco. O tipo de descritor molecular escolhido foi o de fragmento, visto que esse é o mais ideal para se relacionar os valores gerados com as estruturas químicas apresentadas pelas moléculas presentes no conjunto de dados [23]. Todos os descritores gerados, assim como o significado de cada um deles, estão apresentados na Tabela 1.

Tabela 1. Conjunto de descritores gerados para as moléculas do banco de dados.

Número	Código	Significado
001	nCIC	Número de ciclos

002	Rbrid	Número de pontes de ciclos
003	MCD	Grau de ciclização molecular
004	RFD	Densidade de fusão de anéis
005	NRS	Número de sistemas de anéis
006	NNRS	Número de sistemas de anéis normalizado
007	NSpiro	Número de átomos espiro
008	NBridgeHead	Número de átomos cabeça de ponte
009	nR04	Número de anéis com quatro membros
010	nR06	Número de anéis com seis membros
011	nR07	Número de anéis com sete membros
012	nR08	Número de anéis com oito membros
013	nR09	Número de anéis com nove membros
014	nR10	Número de anéis com dez membros
015	nR11	Número de anéis com onze membros
016	nR12	Número de anéis com doze membros
017	AP	Proporção aromática
018	D/Dtr03	Índice de desvio de anel de ordem três
019	D/Dtr06	Índice de desvio de anel de ordem seis
020	D/Dtr07	Índice de desvio de anel de ordem sete
021	D/Dtr09	Índice de desvio de anel de ordem nove
022	D/Dtr10	Índice de desvio de anel de ordem dez
023	D/Dtr11	Índice de desvio de anel de ordem onze
024	D/Dtr12	Índice de desvio de anel de ordem doze
025	nCp	Número de carbonos primários (sp^3)
026	nCs	Número de carbonos secundários (sp^3)
027	nCt	Número de carbonos terciários (sp^3)
028	nCq	Número de carbonos quaternários (sp^3)
029	nCbH	Número de benzenos não-substituídos (sp^2)
030	nCconj	Número de carbonos conjugados não aromáticos (sp^2)
031	nR=Cp	Número de carbonos primários (sp^2)
032	nR=Ct	Número de carbonos terciários alifáticos (sp^2)
033	nRCOOH	Número de ácidos carboxílicos (alifáticos)

034	nRCOOR	Número de ésteres (alifáticos)
035	nArCOOR	Número de ésteres (aromáticos)
036	nRCONHR	Número de amidas secundárias (alifáticas)
037	nRCHO	Número de aldeídos (alifáticos)
038	nArCO	Número de cetonas (aromáticas)
039	nRC=N	Número de iminas (alifáticas)
040	nRNH ₂	Número de aminas primárias (alifáticas)
041	nArNH ₂	Número de aminas primárias (aromáticas)
042	nRNHR	Número de aminas secundárias (alifáticas)
043	nRNR ₂	Número de aminas terciárias (alifáticas)
044	nArNR ₂	Número de aminas terciárias (aromáticas)
045	nN ⁺	Número de nitrogênios com carga positiva
046	nRNHO	Número de hidroxilaminas (alifáticas)
047	nN(CO) ₂	Número de imidas (alifáticas)
048	nArOH	Número de hidroxilas aromáticas
049	nOHp	Número de álcoois primários
050	nOHs	Número de álcoois secundários
051	nOHt	Número de álcoois terciários
052	nSO ₃	Número de sulfonatos
053	nSO ₄	Número de sulfatos
054	nCHRX ₂	Número de CHRX ₂
055	nCXr	Número de heteroátomos em anéis de carbonos (sp ³)
056	nOxiranes	Número de oxiranos
057	nOxetanes	Número de oxetanos
058	nPyrrolidines	Número de pirrolidinas
059	nPyridines	Número de piridinas
060	nHDon	Número de átomos doadores de ligação de hidrogênio (N e O)
061	nHAcc	Número de átomos aceptores de ligação de hidrogênio (N, O, F)
062	C-001	CH ₃ R / CH ₄
063	C-002	CH ₂ R ₂
064	C-003	CHR ₃

065	C-005	CH_3X
066	C-006	CH_2RX
067	C-007	CH_2X_2
068	C-008	CHR_2X
069	C-009	CHRX_2
070	C-011	CR_3X
071	C-012	CR_2X_2
072	C-016	$=\text{CHR}$
073	C-018	$=\text{CHX}$
074	C-019	$=\text{CRX}$
075	C-020	$=\text{CX}_2$
076	C-025	$\text{R}-\text{CR}-\text{R}$
077	C-026	$\text{R}-\text{CX}-\text{R}$
078	C-028	$\text{R}-\text{CR}-\text{X}$
079	C-033	$\text{R}-\text{CH}..\text{X}$
080	C-034	$\text{R}-\text{CR}..\text{X}$
081	C-036	$\text{Al}-\text{CH}=\text{X}$
082	C-038	$\text{Al}-\text{C}(=\text{X})-\text{Al}$
083	C-040	$\text{R}-\text{C}(=\text{X})-\text{X} / \text{R}-\text{C}\equiv\text{X} / \text{X}=\text{C}=\text{X}$
084	H-046	H ligado a $\text{C0}(\text{sp}^3)$ no X ligado ao próximo C
085	H-047	H ligado a $\text{C1}(\text{sp}^3)/\text{C0}(\text{sp}^2)$
086	H-048	H ligado a $\text{C2}(\text{sp}^3)/\text{C1}(\text{sp}^2)/\text{C0}(\text{sp})$
087	H-049	H ligado a $\text{C3}(\text{sp}^3)/\text{C2}(\text{sp}^2)/\text{C3}(\text{sp}^2)/\text{C3}(\text{sp})$
088	H-051	H ligado a alfa-C
089	H-052	H ligado a $\text{C0}(\text{sp}^3)$ com 1X ligado ao próximo C
090	H-053	H ligado a $\text{C0}(\text{sp}^3)$ com 2X ligado ao próximo C
091	H-054	H ligado a $\text{C0}(\text{sp}^3)$ com 3X ligado ao próximo C
092	O-056	Álcool
093	O-057	Fenol / enol / OH de carboxila
094	O-058	$=\text{O}$
095	O-059	$\text{Al}-\text{O}-\text{Al}$
096	O-060	$\text{Al}-\text{O}-\text{Ar} / \text{Ar}-\text{O}-\text{Ar} / \text{R}..\text{O}..\text{R} / \text{R}-\text{O}-\text{C}=\text{X}$
097	O-062	O^- (carregado negativamente)

098	N-067	Al ₂ -NH
099	N-072	RCO-N< / >N-X=X
100	N-073	Ar ₂ NH / Ar ₃ N / Ar ₂ N-Al / R..N..R
101	N-074	R#N / R=N-
102	N-075	R-N-R / R-N-X
103	Cl-087	Cl ligado a C2(sp ³)
104	S-110	R-SO ₂ -R

4.3 Mapas auto-organizáveis e 5-fold

Empregou-se o SOM Toolbox 2.0 [5,13] para analisar os descritores moleculares previamente selecionados. Este kit de ferramentas MATLAB facilita o desenvolvimento e implementação de mapas auto-organizáveis (SOMs), permitindo sua criação, visualização e análise [5,13]. Inicialmente, o conjunto de dados foi inserido na rede sem quaisquer ajustes preliminares. O conjunto de dados foi então dividido em subgrupos com base nos vetores de peso das regiões do mapa em cada etapa de treinamento. A precisão das previsões para esses subgrupos e para o conjunto de dados geral foi avaliada. Para os modelos mais promissores, o conjunto de dados foi dividido em conjuntos de treinamento e teste para avaliar o desempenho preditivo. As precisões de treinamento e teste foram calculadas como a proporção de amostras classificadas corretamente. Para garantir robustez, a validação cruzada de 5 vezes foi aplicada, no Knime 5.3.0 [14], a cada mapa, dividindo os dados em 80% para treinamento e 20% para teste. Dentro do SOM, regiões específicas foram identificadas que agrupavam moléculas com valores de descritores semelhantes, revelando padrões químicos subjacentes.

4.4 Docking Molecular e Propriedades ADMET

Todas as moléculas foram modeladas usando o software MarvinSketch v.23.14 [69]. A otimização foi realizada utilizando o método MMFF implementado no Spartan '14' [70]. As estruturas cristalográficas de quatro alvos envolvidos com a leishmaniose foram obtidas no Protein Data Bank (PDB), e estão apresentadas na Tabela 2.

Tabela 2. Alvos e ligantes usados no docking molecular.

Alvo	PDB ID e Referência	Resolução	Ligante	Espécie
Ornitina descarboxilase	2O00 [71]	1,90 Å	Fosfato de piridoxal	<i>Leishmania donovani</i>
14-alfa-desmetilase	3L4D [72]	1,35 Å	Fluconazol	<i>Leishmania infantum</i>
PnC1	3R2J [73]	2,68 Å	Niacina	<i>Leishmania infantum</i>
Piridoxal quinase	6K91 [74]	2,00 Å	Piridoxina	<i>Leishmania donovani</i>

O software Molegro Virtual Docker 6.0 [75] foi empregado para realizar o docking molecular. Todas as moléculas de água e cofatores foram excluídas do espaço de trabalho. Antes do docking molecular, uma etapa de redocking foi conduzida para avaliação da acurácia e confiabilidade dos resultados do docking, e o Desvio Quadrático Médio da Raiz (do inglês, RMSD) foi calculado. O RMSD é uma etapa crucial para verificar se o algoritmo foi capaz de gerar a mesma pose do arquivo PDB e valores ≤ 2 Å são considerados aceitáveis.

A simulação foi conduzida usando parâmetros padrão. A função MolDock Score foi usada para avaliar as poses dos ligantes, considerando contribuições de energia interna, ligações de hidrogênio e energia torsional. Dez rodadas de docking independentes foram realizadas usando o algoritmo MolDock SE, e as dez poses principais foram retidas. Uma grade com um raio de 15 Å e uma resolução de 0,30 Å foi gerada centrada nas posições dos ligantes cristalográficos das proteínas selecionadas. As poses docadas foram subsequentemente analisadas usando o Discovery Studio Visualizer [76].

Após a realização do docking, as moléculas mais promissoras foram selecionadas e tiveram suas propriedades farmacocinéticas preditas através da ferramenta web SwissADME [89]. Através desse software é possível prever propriedades físico-químicas, farmacocinéticas e de biodisponibilidade [89]. Essas propriedades estão altamente correlacionadas a possível toxicidade de um princípio

ativo, de modo que as mais importantes foram selecionadas, comparando assim as propriedades apresentadas pelas moléculas bem avaliadas no docking com as propriedades apresentadas pela pentamidina, um fármaco presente no tratamento padrão da leishmaniose [90, 108].

CAPÍTULO 5: RESULTADOS E DISCUSSÕES

5. RESULTADOS E DISCUSSÕES

5.1 Banco de Dados

O banco de dados é composto por alcaloides isolados de espécies da família Fabaceae, compreendendo 650 metabólitos secundários e 747 ocorrências botânicas (número de espécies das quais um composto foi isolado), distribuídos entre 129 espécies e 33 gêneros diferentes. Os quatro gêneros com o maior número de estruturas químicas são *Erythrina*, *Sophora*, *Erythrophleum* e *Lupinus*. A distribuição de alcaloides entre esses quatro gêneros é mostrada na Tabela 3. Todos os dados estarão disponíveis na ferramenta web Sistemax (<https://sistemax.ufpb.br/>).

Tabela 3. Distribuição das moléculas encontradas, classificadas por gênero, incluindo dados botânicos e ocorrências químicas.

Gêneros	Espécies	Ocorrências	Alcaloides
<i>Erythrina</i>	36	336	269
<i>Sophora</i>	11	119	105
<i>Erythrophleum</i>	7	68	68
<i>Lupinus</i>	11	52	41
Outros	64	172	167
Total	129	747	650

5.2 Mapas auto-organizáveis e descritores moleculares aplicados à quimiotaxonomia de alcaloides de Fabaceae.

Descritores moleculares são medidas quantitativas obtidas pela aplicação de funções matemáticas a representações específicas de estruturas químicas. Esses descritores capturam informações sobre as propriedades físicas, químicas e eletrônicas das moléculas [4]. Os valores calculados dos descritores foram usados como dados de entrada no software SOM Toolbox 2.0 [5]. Para a construção do mapa auto-organizável, foram selecionadas apenas as ocorrências botânicas dos gêneros com o maior número de estruturas químicas, nomeadamente *Erythrina* e *Sophora*,

totalizando 455 ocorrências e 374 alcaloides (Tabela 3). Um total de 104 descritores moleculares diferentes foram gerados para os metabólitos secundários correspondentes às ocorrências, utilizando o software alvaDesc 2.0 [3]. Descritores de fragmentos foram selecionados porque são os ideais para relacionar os valores gerados às estruturas químicas das moléculas no conjunto de dados [23].

O mapa auto-organizável construído a partir dos descritores moleculares gerados a partir dos alcaloides de *Erythrina* e *Sophora* mostra a taxa de acerto total de 98,5% (Tabela 4). O SOM (Figura 7) mostrou uma clara separação entre as ocorrências químicas dos dois gêneros, *Erythrina* e *Sophora*, o que corrobora com a análise filogenética realizada por Zhao et al [6]. Analisando o mapa, pode-se observar que as ocorrências derivadas do gênero *Erythrina* ocupam um número substancialmente maior de neurônios, o que está de acordo com a porcentagem presente no conjunto de dados e apresentada na Tabela 4, uma vez que o número de ocorrências químicas apresentadas pelo gênero *Erythrina* representa 73,8% do total de dados utilizados para a construção do mapa auto-organizável. No entanto, também é importante destacar que o maior número de neurônios ocupados pode estar relacionado à maior diversidade química estrutural apresentada pelas ocorrências do gênero *Erythrina*.

Tabela 4. Resultados do mapa auto-organizável com os valores das ocorrências e o número de acertos corretos para os gêneros *Erythrina* e *Sophora* da família Fabaceae.

Gêneros	Ocorrências	% de Ocorrências	No. de Acertos	% de Acertos
<i>Erythrina</i>	336	73.8	334	99.4
<i>Sophora</i>	119	26.2	114	98.8
Total	455	100	448	98.5

O desempenho preditivo do SOM, que foi baseado em cinco diferentes conjuntos de treinamento e teste, derivados do banco de dados usado para a construção do mapa, é apresentado na Tabela 5. O gênero *Erythrina* mostra os maiores valores médios de taxa de correspondência para conjuntos de treinamento (98,9%) e testes (97,9%), enquanto *Sophora* mostrou 96,9% e 94,9% para conjuntos

de treinamento e testes, respectivamente, que são resultados semelhantes aos obtidos para o conjunto de dados completo, 99,4% para *Erythrina* e 98,8% para *Sophora* (Tabela 4).

Tabela 5. Resumo dos resultados de correspondência de treinamento e teste em porcentagem de *Erythrina* e *Sophora*.

Gêneros	Treino 1	Treino 2	Treino 3	Treino 4	Treino 5	Média
<i>Erythrina</i>	99.6	98.1	98.8	98.5	99.6	98.9
<i>Sophora</i>	95.8	96.8	95.7	97.9	93.6	96.9
Total	98.6	97.8	98.0	98.3	98.0	98.1
Genus	Test Set 1	Test Set 2	Test Set 3	Test Set 4	Test Set 5	Average
<i>Erythrina</i>	98.5	98.5	100	98.5	94.0	97.9
<i>Sophora</i>	95.6	95.8	87.5	95.8	100	94.9
Total	97.8	97.8	96.7	97.8	95.6	97.1

O SOM mostra as ocorrências químicas do gênero *Erythrina* em vermelho e *Sophora* em verde, os neurônios pretos representam regiões no mapa que não foram preenchidas pelas ocorrências. A U-matrix exibe a distância entre as unidades de mapa vizinhas, o que significa que quanto maior o valor correspondente, maior a distância entre os clusters. Inversamente, quanto menor o valor correspondente, menor a diferença entre os clusters. Para a análise de componentes principais (PCA), que é medida usando os maiores vetores e autovalores, as unidades vizinhas presentes no mapa são interconectadas por linhas para facilitar a visualização dos dados e, conseqüentemente, ao usar duas variáveis principais para sua construção, é possível ver as distâncias entre os componentes do mapa em 2D. Os dados são apresentados na Figura 7.

A Figura 8 exibe os descritores moleculares mais importantes que diferenciam os compostos encontrados em *Erythrina* e *Sophora*. Os três descritores exibiram padrões semelhantes, com valores mais altos geralmente associados a alcaloides de *Sophora* e valores mais baixos a alcaloides de *Erythrina*.

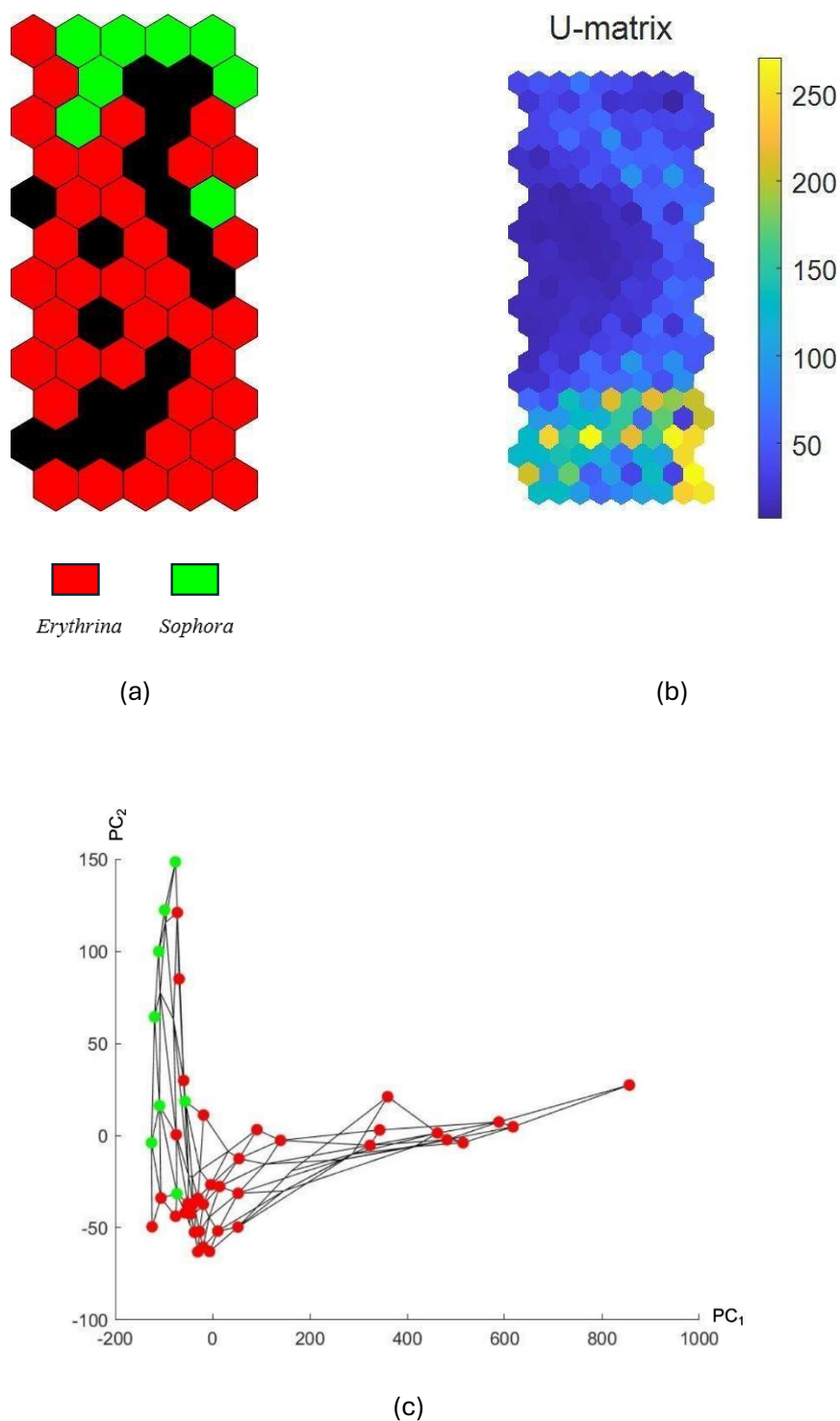


Figura 7. Mapa auto-organizável obtido pela classificação dos gêneros *Erythrina* e *Sophora*: (a) SOM dos descritores moleculares; (b) U-Matrix; (c) Projeção PCA do SOM.

É possível observar que o gênero *Sophora* apresenta um número menor de ocorrências botânicas (26,2% do banco de dados) para a construção do SOM e, consequentemente, do PCA. No entanto, apresentou uma ótima taxa de acerto, assim

como o gênero *Erythrina*, além de que diferenças estruturais são perceptíveis, já que o PCA apresentou regiões bem definidas tanto para as ocorrências de *Erythrina* quanto de *Sophora*. O PCA realizado apresentou uma variância explicada de 95,55%, ou seja, usando apenas duas variáveis é possível visualizar quase toda a variação demonstrada pelos dados presentes no SOM.

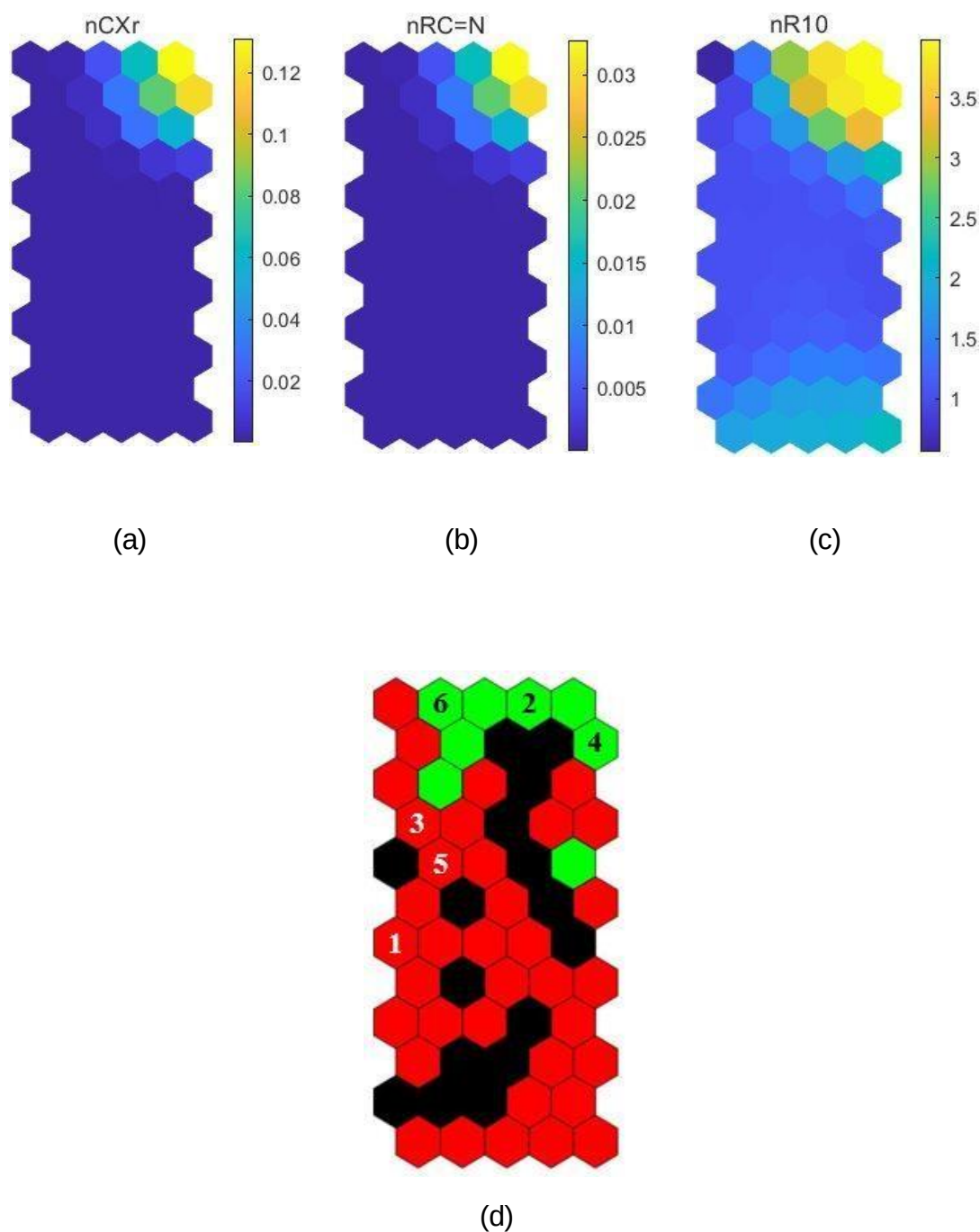
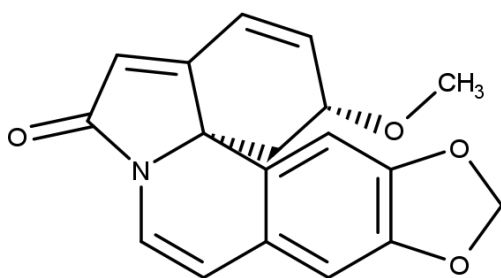


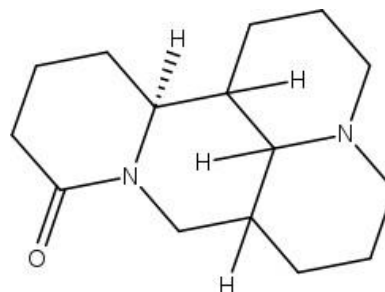
Figura 8. Mapa auto-organizável obtido pela classificação dos gêneros *Erythrina* e *Sophora*: (a) Descritor molecular nCXr; (b) Descritor molecular nRC=N; (c) Descritor

molecular nR10; (d) Posição dos compostos 1-6 no SOM.

O descritor nCXr está relacionado à quantidade de grupos funcionais ou átomos diferentes de carbono, ligados a um átomo de carbono pertencente a um anel hibridizado sp^3 . Consequentemente, é evidente que o alcaloide 1 (Figura 9), localizado na região descrita pela Figura 8d e comumente conhecido como cristamidina [7], possui a menor quantidade de nitrogênios ligados a carbonos saturados do que o alcaloide 2 (Figura 9), conhecido como matrina [8], que também está localizado na região descrita pela Figura 8d. Esse fato pode estar diretamente ligado aos valores indicados pelo descritor no mapa, sugerindo que os alcaloides de *Sophora* analisados, em geral, exibem uma maior quantidade de anéis formados por carbonos com apenas ligações simples.



1 (Cristamidina)
Erythrina crista-galli



2 (Matrina)
Sophora alopecuroides

Figura 9. Alcaloides que exibem a diferença mostrada pelo descritor nCXr no SOM.

Para o descritor nCR=N, que está relacionado ao número de iminas não aromáticas, pode-se observar que o alcaloide 3 (Figura 10), localizado na região descrita pela Figura 8d e conhecido como erisodina [7], não possui nenhum grupo funcional imina em sua estrutura. Esse padrão é geralmente repetido para alcaloides pertencentes ao gênero *Erythrina* que foram analisados. Em contraste, a presença de uma imina não aromática na estrutura do alcaloide 4 (Figura 10), que está na região apresentada na Figura 8d e comumente conhecido como flasevina I [9], é claramente observável.

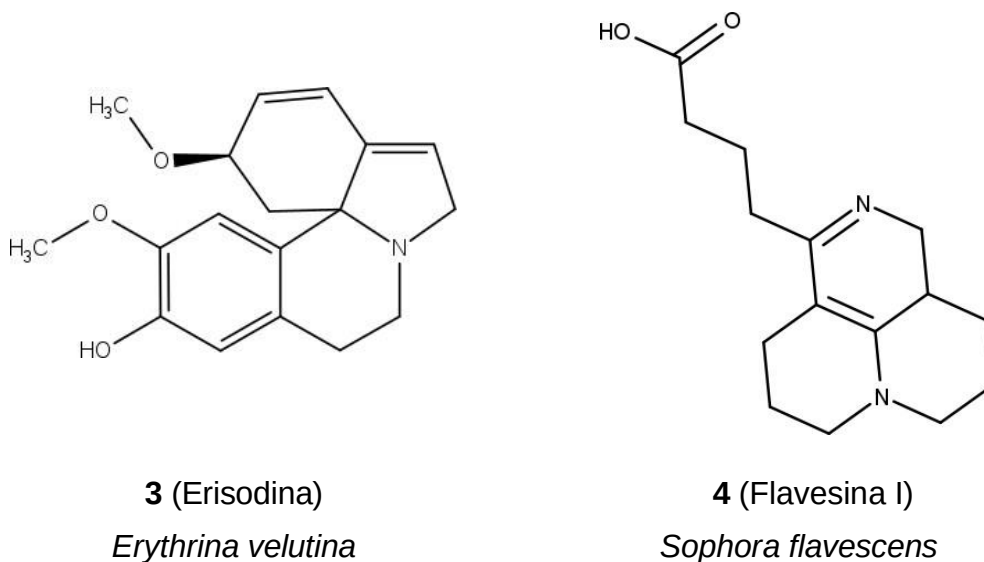


Figura 10. Alcaloides que exibem a diferença mostrada pelo descritor nRC=N no SOM.

Em relação ao descritor nR10, que está relacionado ao número de anéis de dez membros, pode-se observar que o alcaloide 6 (Figura 11), conhecido como alopecurina A [8] e localizado na região descrita pela Figura 8d, possui um anel de dez membros em sua estrutura, enquanto o alcaloide 5 (Figura 11), conhecido como erisotrina [7] e localizado conforme mostrado na Figura 8d, não possui. A presença de um anel de dez membros é um padrão frequentemente observado no conjunto de dados analisado para ocorrências do gênero *Sophora*.

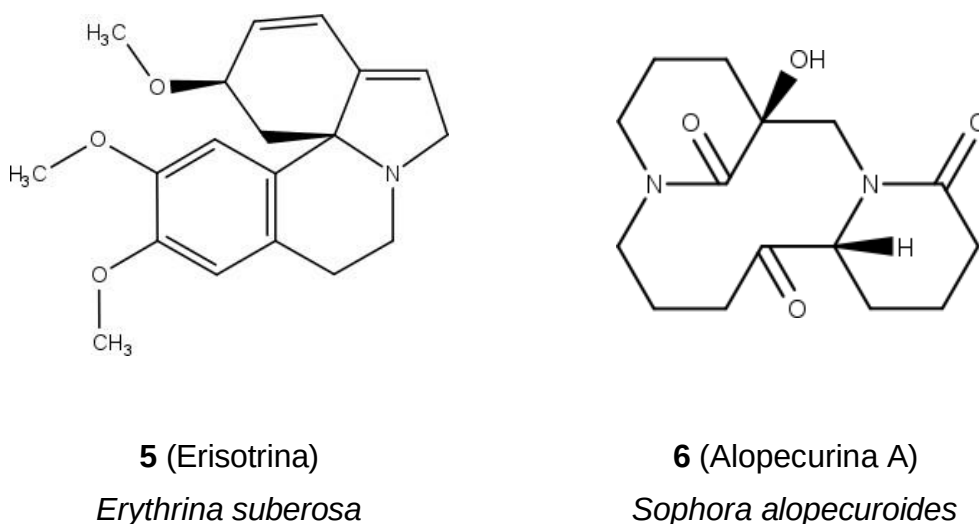


Figura 11. Alcaloides que exibem a diferença mostrada pelo descritor nR10 no SOM.

5.3 Biossíntese de alcaloides de *Erythrina* e *Sophora*

As diferenças estruturais observadas entre os alcaloides podem ser atribuídas às distintas vias biossintéticas envolvidas em sua formação dentro de cada gênero. A presença de anéis hibridizados sp^3 pode estar relacionada à via que se inicia a partir da L-lisina (L-Lysine), levando à formação de (-)-esparteína ((-)-Sparteine), conforme descrito por Huang et al. [10]. Essa via biossintética está envolvida na formação estrutural dos alcaloides quinolizidínicos de *Sophora*. A outra via descrita por Huang et al. [10], promovendo a formação de piperideína, pode estar ligada à formação de iminas não aromáticas observadas em alcaloides de *Sophora*. As informações listadas estão apresentadas na Figura 1.

Inversamente, as espécies de *Erythrina* seguem uma via biossintética distinta, começando com os aminoácidos fenilalanina e tirosina para produzir alcaloides isoquinolínicos, conforme descrito por Chacon et al [11]. Essa diferença pode estar relacionada à formação de alcaloides com um maior grau de insaturação nos anéis de suas estruturas centrais, reforçando assim a diferença produzida pelo descritor nCXr. As informações listadas estão apresentadas na Figura 2.

5.4 Atividade biológica dos alcaloides de *Erythrina* e *Sophora*.

Os alcaloides de *Sophora* demonstraram potencial promissor em várias áreas farmacológicas, incluindo atividades antitumorais, antimicrobianas e anti-inflamatórias, conforme destacado por Wang et al. [8]. Especificamente, os alcaloides 2 (Figura 9) e 6 (Figura 11) mostraram efeitos antitumorais e anti-inflamatórios notáveis, o que pode ser atribuído à sua estrutura central quinolizidínica compartilhada. A proximidade dessas ocorrências no SOM, conforme mostrado na Figura 8d, pode estar relacionada às semelhanças estruturais e de atividade apresentadas pelas moléculas.

Da mesma forma, os alcaloides de *Erythrina* foram relatados como possuidores de uma ampla gama de atividades biológicas, como propriedades antibacterianas, antifúngicas, antiprotozoárias e anticonvulsivantes, conforme detalhado por Rambo et al. [7]. Entre os alcaloides de *Erythrina*, o composto 1 (Figura 9) exibe atividade antiprotozoária promissora, enquanto os compostos 3 (Figura 10) e 5 (Figura 11) demonstraram efeitos ansiolíticos. As semelhanças estruturais e de atividade apresentadas pelos compostos 3 e 5 podem estar relacionadas à proximidade de suas

posições no SOM (Figura 4d), enquanto a diferença apresentada pelo composto 1 pode estar relacionada à sua maior distância no mapa (Figura 8d) quando comparado às posições dos alcaloides 3 e 5.

5.5 Docking Molecular e Propriedades ADMET

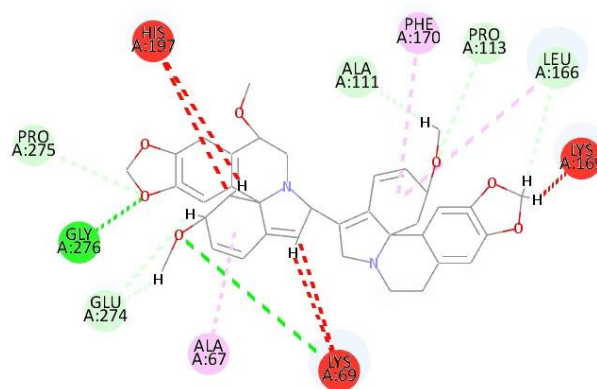
Os resultados do docking molecular para ornitina descarboxilase estão apresentados na Tabela 6. O MolDock Score fornece uma estimativa da energia de ligação para complexos formados entre os ligantes e os alvos, em que valores mais negativos sugerem complexos mais estáveis [5].

Tabela 6. Cinco melhores moléculas obtidas no docking com a ornitina descarboxilase

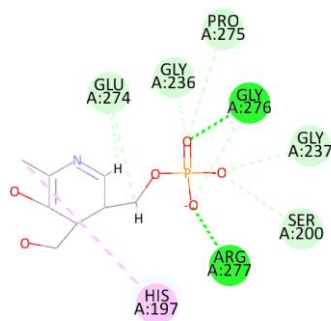
Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
Ornitina descarboxilase	2O00	Eritrivarina A5	-151,03	0.14 Å
		Saikachinosidio B	-145,68	
		Indol-3-il-metil glicosinolato	-144,83	
		Saikachinosidio C	-144,73	
		Eritrivarina A8	-135,46	
		Fosfato de piridoxal	-94,48	

As análises revelam que as cinco moléculas destacadas na Tabela 6 formaram complexos mais estáveis com a ornitina decarboxilase quando comparados com o complexo formado com o fosfato de piridoxal [71]. Isso sugere que essas moléculas,

com destaque para a eritrivarina A5, com escore de -151,03, possuem melhores afinidades de ligação com a proteína alvo quando comparadas com o ligante cristalografado, que teve escore de -94,48.



Eritrivarina A5



Fosfato de piridoxal

Legenda (inglês) [76]:

van der Waals	Conventional Hydrogen Bond	Unfavorable Negative-Negative	Salt Bridge	Pi-Anion
Carbon Hydrogen Bond	Pi-Pi Stacked	Halogen (Cl, Br, I)	Pi-Sulfur	
Pi-Alkyl	Pi-Sigma	Pi-Anion		

Figura 12. Interações com os resíduos de aminoácidos da ornitina decarboxilase

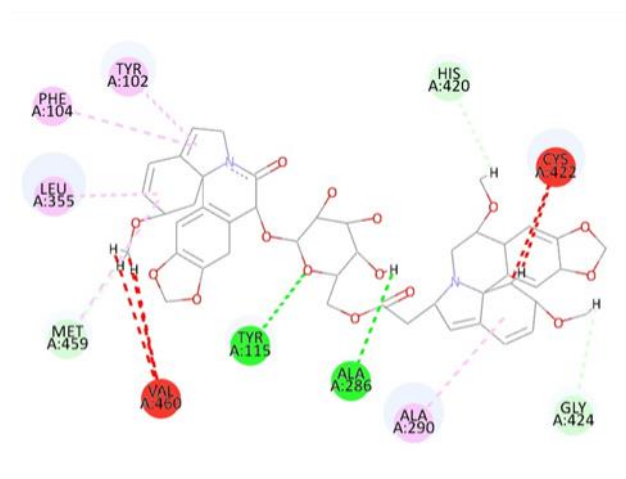
Apesar de apresentar uma quantidade superior de interações desfavoráveis na formação do complexo, a eritrivarina A5 apresentou um valor de escore menor, quando comparada o fosfato de piridoxal. Esse fato pode estar relacionado a maior

quantidade de interações de hidrogênio com carbono estabelecidas pelo complexo entre a eritrivarina A5 e a ornitina descarboxilase, a exemplo dos resíduos Ala111 (alanina) e Leu166 (leucina). As interações desfavoráveis estabelecidas com o resíduo Lys69 (lisina) são críticas pois o mecanismo de inibição se inicia através da quebra das ligações com esse resíduo [71, 92]. As interações com os resíduos das outras quatro moléculas mais bem ranqueadas, e as dez moléculas com melhores valores de escore estão apresentadas nos Apêndice I e II, respectivamente.

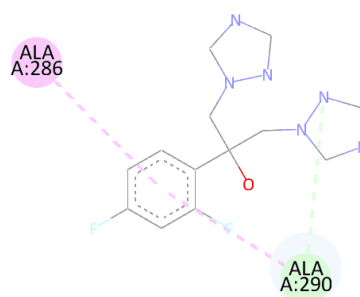
Tabela 7. Cinco melhores moléculas obtidas no docking com a 14-alfa-desmetilase

Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
Urato Oxidase	3LD4	Eritrivarina D	-194,59	0.22 Å
		Eritrivarina C	-190,96	
		Eritrivarina B	-163,10	
		Eritrivarina U	-156,93	
		Eritrivarina F	-156,35	
		Fluconazol	-73,23	

As simulações de docking revelam que as cinco moléculas destacadas na Tabela 7 formaram complexos mais estáveis com a urato oxidase quando comparados com o complexo formado com o fluconazol [72]. Isso sugere que essas moléculas, com destaque para a eritrivarina D, com escore de -194,59, possuem melhores afinidades de ligação com a proteína alvo quando comparadas com o ligante cristalografado, que teve escore de -73,23. Pode-se destacar também que para essa proteína as moléculas ranqueadas com melhor escore são muito semelhantes em sua estrutura, o que pode sugerir que o cerne de suas estruturas químicas apresenta boa afinidade com a urato oxidase.



Eritrivarina D



Fluconazol

Legenda (inglês) [76]:

van der Waals	Conventional Hydrogen Bond	Unfavorable Negative-Negative	Salt Bridge	Pi-Anion
Carbon Hydrogen Bond	Pi-Pi Stacked	Halogen (Cl, Br, I)	Pi-Sulfur	
Pi-Alkyl	Pi-Sigma	Pi-Anion		

Figura 13. Interações com os resíduos de aminoácidos 14-alfa-desmetilase

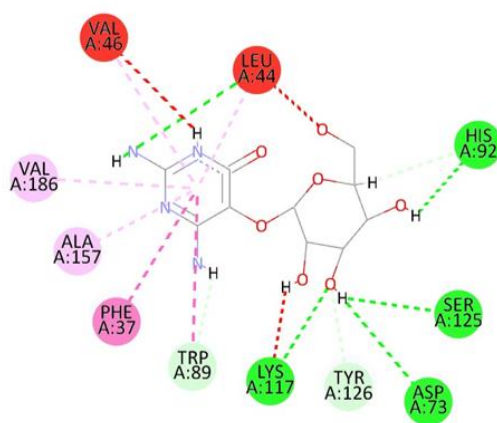
Apesar de apresentar uma quantidade superior de interações desfavoráveis, mais especificamente com os resíduos Val460 (valina) e Cys422 (cisteína), a eritrivarina D apresentou valor de escore muito menor que o fluconazol, o que pode estar relacionado com a quantidade de interações de ligação de hidrogênio realizadas como uma quantidade superior de diferentes resíduos, mais especificamente Tyr115

(tirosina) e Ala286 (alanina), da urato oxidase. A eritrivarina D também apresentou interações de hidrogênio com carbono, nos aminoácidos Met459 (metionina) e His420 (histidina), o que contribuí para estabilização do complexo, evidenciando assim interações que não estão presentes no complexo formado com a fluconazol. Apesar de os dois complexos terem apresentado interações com resíduos idênticos, a exemplo da Ala290 (Alanina), os dois complexos não apresentaram interações consideradas críticas para a inibição da enzima [72]. As interações com os resíduos das outras quatro moléculas mais bem ranqueadas, e as dez moléculas com melhores valores de escore estão apresentadas nos Apêndices IV e III, respectivamente.

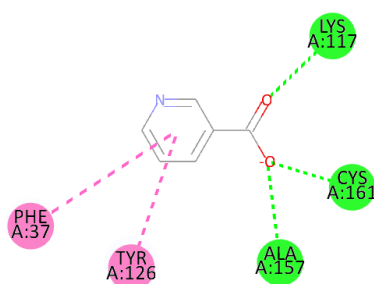
Tabela 8. Cinco melhores moléculas obtidas no docking com a PnC1

Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
PnC1	3R2J	Vicina	-124,46	0.28 Å
		Monocrotalina	-117,69	
		Erisopina	-113,76	
		N,N-Bis(2-(1 <i>H</i> -indol-3-il) etil) oxalamida	-113,51	
		Convicina	-110,47	
		Niacina	-54,32	

Os valores de escore obtidos no docking revelam que as cinco moléculas destacadas na Tabela 8 também formaram complexos mais estáveis, mas nesse caso com a PnC1, quando comparados com o complexo formado com a niacina [73]. Isso sugere que essas moléculas, destaque para a vicina com escore de -124,46, possuem melhores afinidades de ligação com a proteína alvo quando comparadas com o ligante cristalografado, que teve escore de -54,32.



Vicina



Niacina

Legenda (inglês) [76]:

van der Waals	Conventional Hydrogen Bond	Unfavorable Negative-Negative	Salt Bridge	Pi-Anion
Carbon Hydrogen Bond	Pi-Pi Stacked	Halogen (Cl, Br, I)	Pi-Sulfur	
Pi-Alkyl	Pi-Sigma	Pi-Anion		

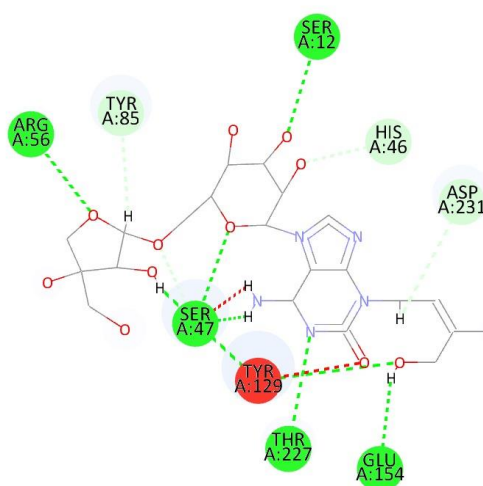
Figura 14. Interações com os resíduos de aminoácidos da PnC1

O menor valor de escore apresentado pela vicina pode estar relacionado a quantidade superior de ligações de hidrogênio apresentadas na formação de seu complexo com PnC1, quando em comparação com o complexo formado com a niacina, podendo-se destacar os resíduos Lys117 (lisina), Ser125 (serina), Asp73 (ácido aspártico), His92 (histidina). É importante destacar que a vicina apresentou interações críticas para inibição da enzima, mais especificamente a interação pi-alquila e a ligação de hidrogênio com os resíduos Val186 (valina) e His92 (histidina) [92]. As interações com os resíduos das outras quatro moléculas mais bem ranqueadas, e as dez moléculas com melhores valores de escore estão apresentadas nos Apêndices VI e V, respectivamente.

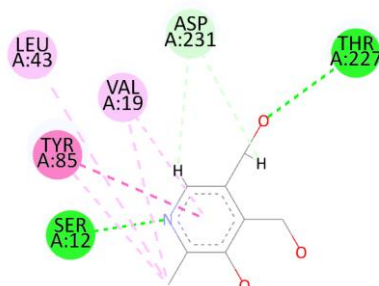
Tabela 9. Cinco melhores moléculas obtidas no docking com a piridoxal quinase

Alvo	PDB ID	Ligante	MolDock Score	RMSD
Piridoxal quinase	6k91	Saikachinsidio B	-174,61	0.10 Å
		Locustosidio B	-170,82	
		Prosojulinina	-166,40	
		Indol-3-il-metil glicosinolato	-162,92	
		Juliprosinina	-162,63	
		Piridoxina	-69,69	

Para a piridoxal quinase, os valores de escore obtidos no docking demonstraram que as cinco moléculas destacadas na Tabela 9 também formaram complexos mais estáveis, quando comparados com o complexo formado com a piridoxina [74]. Isso sugere que essas moléculas, destacando-se a saikachinosidio B que obteve escore de -174,61, possuem melhores afinidades de ligação com a proteína alvo quando comparadas com o ligante cristalografado, que teve escore de -69,69. Pode-se concluir que a molécula mais promissora da simulação de docagem apresentou um escore em mais de cem unidades inferior a piridoxina.



Saikachinsidio B



Piridoxina

Legenda (inglês) [76]:

van der Waals	Conventional Hydrogen Bond	Unfavorable Negative-Negative	Salt Bridge	Pi-Anion
Carbon Hydrogen Bond	Pi-Pi Stacked	Halogen (Cl, Br, I)	Pi-Sulfur	
Pi-Alkyl	Pi-Sigma	Pi-Anion		

Figura 15. Interações com os resíduos de aminoácidos da piridoxal quinase.

O menor valor de escore apresentado pela saikachinsidio B pode estar relacionado a quantidade superior de ligações de hidrogênio apresentadas na formação de seu complexo com piridoxal quinase, quando em comparação com o complexo formado com a piridoxina, podendo-se destacar os resíduos Arg56 (arginina), Ser12 (serina), Ser47 (serina), Thr227 (treonina). Além disso a formação do complexo com a saikachinsidio B apresentou o estabelecimento de interações

críticas favoráveis com os resíduos Ser47 (serina), Ser12 (serina), Asp231 (ácido aspártico) [104]. As interações com os resíduos das outras quatro moléculas mais bem ranqueadas, e as dez moléculas com melhores valores de escore estão apresentadas nos Apêndices VIII e VII, respectivamente.

As substâncias com os melhores resultados no docking molecular foram avaliadas segundo suas propriedades físico-químicas, farmacocinéticas e de biodisponibilidade. Essas moléculas tiveram suas propriedades farmacocinéticas comparadas com as da pentamidina [108]. A predição das propriedades farmacocinéticas é uma importante etapa no estudo de uma nova droga e cerca de 25% da maioria dos fármacos testados não chega às fases clínicas devido a problemas farmacocinéticos e tóxicos [93].

Dentre os fatores analisados, a lipofilicidade é capaz de influenciar o desenvolvimento de uma nova droga, desde a sua formulação, à distribuição e absorção no indivíduo [93, 94]. O principal método de análise da lipofilicidade é por meio do logaritmo do coeficiente octanol/água (LogP). A plataforma utilizada (SwissADME) para a estimativa do LogP utiliza diferentes preditores, como XLOGP3, MLOGP, SILICOS-IT, iLOGP e o WLOGP, e fornece, portanto, uma estimativa mais robusta por diferentes métodos. A partir da média aritmética desses preditores, pode-se calcular o consenso.

Estima-se que o intervalo ideal para o coeficiente de partição logP esteja entre 1 e 3, uma vez que valores inferiores a 1 indicam baixa lipofilicidade, o que compromete a permeabilidade através dos enterócitos, enquanto valores superiores a 3 refletem elevada lipofilicidade, resultando em baixa solubilidade nos fluidos gastrointestinais e, conseqüentemente, em absorção limitada [94]. Observou-se que, além da pentamidina, os compostos erisopina, N,N-Bis(2-(1H-indol-3-il)etil) oxalamida, prosojuliflavona e erithrivarinas D e F apresentaram logP dentro desse intervalo (Tabela 10), sugerindo que essas moléculas possuem maior potencial para serem eficientemente absorvidas por via oral.

A área de superfície topológica polar (Topological Polar Surface Area, TPSA), que representa a soma das áreas de todos os grupos polares de uma molécula, constitui um parâmetro determinante no desenvolvimento de novas drogas [95]. Prediz-se que compostos com TPSA inferior a 140 Å² apresentam maior probabilidade de exibir boa absorção passiva no trato gastrointestinal [95]. Nesse contexto, verificou-se que, além da pentamidina, os compostos erisopina, juliprosinina, monocrotalina,

N,N-Bis(2-(1H-indol-3-il)etil) oxalamida e as erithrivarinas U, A5 e A8 apresentaram TPSA abaixo desse limite (Tabela 10), indicando um potencial favorável para absorção oral.

Tabela 10. Propriedades de lipofilicidade e área de superfície polar topológica para as moléculas analisadas, destacando aquelas com bons resultados em um parâmetro (negrito) ou ambos (vermelho).

Molécula	iLOGP	XLOGP	WLOGP	MLOGP	SILICOS-IT	Consenso Log P	tPAS (Å²)
Pentamidina	2.24	2.56	2.88	2.8	3.11	2.72	118.2
Convicina	-1.58	-3.39	-4.17	-2.9	-2.15	-2.84	191.12
Erisopina	2.45	1.49	1.58	1.66	1.65	1.77	52.93
Erithrivarina C	4.82	-0.36	-0.2	-0.65	0.2	0.76	204.61
Erithrivarina D	4.29	0.17	0.45	-0.49	0.77	1.04	193.61
Erithrivarina F	5.89	1.69	2.87	1.25	3.14	2.97	159.55
Erithrivarina U	4.77	3	3.39	2.34	3.31	3.36	78.93
Erithrivarina A5	5.44	2.78	3.36	2.66	3.68	3.58	71.09
Erithrivarina A8	4.96	3.21	3.29	2.89	4.12	3.69	78.93
Indol-3-il-metil glicosinolato	0.37	-0.09	0.46	-1.79	-1.18	-0.45	215.58
Juliprosinina	4.59	10.77	7.13	5.04	9.65	7.43	68.4
Locustosídio B	1.6	-3.44	-3.74	-3.41	-3.35	-2.47	227.8
Monocrotalina	2.33	-0.73	-0.77	0.24	0.24	0.26	96.3
N,N-Bis(2-(1H-indol-3-il)etil) oxalamida	2.71	3.45	2.67	1.43	4.38	2.93	89.78
Prosojuliflavona	2.7	1.5	1.35	-1.3	1.46	1.14	216.58
Saikachinosídio C	0.03	-4.69	-4.77	-4.14	-3.94	-3.5	248.03
Saikachinosídio B	0.03	-4.69	-4.77	-4.14	-3.94	-3.5	248.03
Vicina	-0.31	-3.72	-3.87	-2.95	-3.01	-2.77	197.17

Observa-se que diversos fármacos disponíveis no mercado compartilham características físico-químicas semelhantes, incluindo peso molecular, coeficiente de partição (logP), número de doadores e aceptores de ligações de hidrogênio, entre outros parâmetros. Isso permitiu que diversos pesquisadores desenvolvessem conjuntos de regras para prever se uma determinada substância apresenta propriedades compatíveis com compostos farmacologicamente ativos, criando o termo "druglikeness" [97, 98].

Um dos primeiros modelos utilizados para isso foi a regra dos cinco de Lipinski, que postulou que uma droga possui uma boa absorção ou permeabilidade quando

obedece a quatro regras: a) Peso molecular ≤ 500 da; b) $\log P \leq 5$; c) número de doares de ligações de hidrogênio ≤ 5 ; d) número de aceptores de ligações de hidrogênio ≤ 10 . Duas ou mais violações indica que a molécula não possui características adequadas para ser um fármaco em termos de solubilidade [103].

Ao longo dos anos, diversas adaptações a essas regras foram postuladas, como as propostas por Ghose et al., Ursu et al., Veber et al., 2002 e Muegge et al. [99, 101, 97, 100]. Deve-se ressaltar, no entanto, que apenas uma violação a essas regras não indica necessariamente um problema na biodisponibilidade da droga, de forma que seu estudo como candidato a fármaco não deve ser descartado [102].

A Tabela 11 mostra os resultados de druglikeness utilizando as regras citadas anteriormente. Observa-se que além da pentamidina, apenas erisopina, monocrotalina e N,N-Bis(2-(1H-indol-3-il)etil)oxalamida não apresentam mais de uma violação a essas regras, indicando que elas possuem propriedades físico-químicas mais similares aos fármacos encontrados no mercado atualmente e maior potencial para serem bem absorvidas oralmente.

Tabela 11. Druglikeness das moléculas analisadas, destacando aquelas com no máximo uma violação em todas as regras (negrito)

Molécula	Lipinski #violações	Ghose #violações	Veber #violações	Egan #violações	Muegge #violações
Pentamidina	0	0	0	0	0
Convicina	2	1	1	1	3
Erisopina	0	0	0	0	0
Eritrivarina C	2	3	1	1	4
Eritrivarina D	2	3	1	1	4
Eritrivarina F	2	3	1	1	4
Eritrivarina U	1	3	0	0	2
Eritrivarina A5	1	3	0	0	2
Eritrivarina A8	1	3	0	0	2
Indol-3-il-metil glicosinolato	2	0	1	1	2
Juliprosinina	2	4	1	1	3
Locustosidio B	3	2	1	1	4
Monocrotalina	0	1	0	0	0
N,N-Bis(2-(1H-indol-3-il)etil)oxalamida	0	0	0	0	0
Prosojuliflavona	3	2	1	1	3
Saikachinoside C	3	2	1	1	4

Saikachinosidio B	3	2	1	1	4
Vicina	2	1	1	1	3

Analisando todas as características calculadas, vê-se que as substâncias erisopina, monocrotalina e N,N-Bis(2-(1H-indol-3-il)etil)oxalamida (Figura 16) possuíram bons parâmetros de logP, tPAS e nas regras de druglikeness, indicando que essas moléculas são as que possuem maior potencial para serem novas drogas para o tratamento de leishmaniose baseado nas propriedades farmacocinéticas.

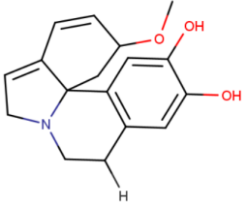
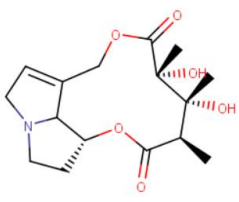
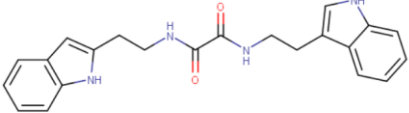
Erisopina	Monocrotalina	N,N-Bis(2-(1H-indol-3-il)etil)oxalamida
		

Figura 16. Alcaloides com melhores resultados farmacocinéticos nos parâmetros analisados.

CAPÍTULO 6: CONCLUSÕES

6. CONCLUSÕES

Este estudo utilizou um rico banco de dados de alcaloides da família Fabaceae, que será disponibilizado na ferramenta web Sistemax (<https://sistemax.ufpb.br>). O banco de dados abrangia 650 metabólitos secundários pertencentes a 129 espécies distintas, resultando em um total de 747 ocorrências botânicas documentadas. Empregando mapas auto-organizáveis (SOMs) e descritores moleculares, foi possível separar efetivamente os gêneros *Sophora* e *Erythrina* com excepcional precisão (superior a 95%). Essas descobertas se alinham com estudos filogenéticos anteriores conduzidos por Zhao et al. [6].

Em essência, os SOMs construídos usando propriedades físico-químicas codificadas de alcaloides servem como uma ferramenta poderosa para identificar estruturas com características específicas. No geral, os alcaloides de *Erythrina* apresentaram uma estrutura com anéis de no máximo seis átomos, com uma maior quantidade de carbonos sp^2 e aminas clássicas. Essa estrutura pode estar relacionada a um potencial farmacológico antibacteriano, antifúngico, anticonvulsivante e antiprotzoário, conforme descrito por Rambo et al. [7]. As espécies de *Sophora*, por outro lado, apresentaram uma estrutura com uma maior quantidade de carbonos sp^3 na formação dos ciclos, anéis com mais de seis átomos e iminas. Essa estrutura pode estar relacionada a um potencial farmacológico antitumoral, antimicrobiano e anti-inflamatório, conforme demonstrado por Wang et al. [8]. Essa abordagem possui grande promessa para aplicações como a identificação de alcaloides com potencial atividade biológica, integrando dados taxonômicos e geográficos.

Além disso, a combinação do docking molecular com a predição de propriedades farmacocinéticas possibilitou selecionar três alcaloides do banco inicial que se apresentam como potenciais fármacos contra leishmaniose, afinal demonstraram escores significantes e interações críticas quando docados com uma das proteínas alvo, ao passo que também apresentaram boas propriedades ADMET, esses foram a erisopina, a monocrotalina e a N,N-Bis(2-(1H-indol-3-il)etil)oxalamida [86, 97, 98]. Contudo é necessário afirmar que esses testes são insuficientes para demonstração de potencial real e consequente realização de testes *in vitro* e *in vivo*, alguns outros testes *in silico*, a exemplo da dinâmica molecular, serão realizados na continuidade do presente estudo.

REFERÊNCIAS

REFERÊNCIAS

1. Chen, Y., Zou, C., Mastalerz, M., Hu, S., Gasaway, C., & Tao, X. (2015). Applications of micro-Fourier transform infrared spectroscopy (FTIR) in the geological sciences – A review. *International Journal of Molecular Sciences*, 16, 30223.
2. de Souza, T. A., Lins, F. S. V., da Silva Lins, J., Alves, A. F., Cibulski, S. P., Brito, T. de A. M., Abreu, L. S., Scotti, L., Scotti, M. T., da Silva, M. S., & Tavares, J. F. (2024). *Asclepiadoideae* subfamily (Apocynaceae): Ethnopharmacology, biological activities and chemophenetics based on pregnane glycosides. *Phytochemistry Reviews*.
3. Mauri, A. (2020). alvaDesc: A tool to calculate and analyze molecular descriptors and fingerprints. In K. Roy (Ed.), *Ecotoxicological QSARs* (pp. 801–820). Humana Press Inc. https://doi.org/10.1007/978-1-0716-0150-1_32
4. Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. In R. Mannhold, H. Kubinyi, & G. Folkers (Eds.). Wiley.
5. Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (n.d.). *SOM toolbox for Matlab*
5. <http://www.cis.hut.fi/projects/somtoolbox/>
6. Zhao, Y., Zhang, R., Jiang, K.-W., Qi, J., Hu, Y., Guo, J., Zhu, R., Zhang, T., Egan, A. N., Yi, T.-S., Huang, C.-H., & Ma, H. (2021). Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant*, 14, 748–773. <https://doi.org/10.1016/j.molp.2021.02.006>
7. Rambo, D. F., Biegelmeyer, R., Toson, N. S. B., Dresch, R., Moreno, P. R. H., & Henriques, A. T. (2019). The genus *Erythrina* L.: A review on its alkaloids, preclinical, and clinical studies. *Phytotherapy Research*, 33(7), 1258–1276. <https://doi.org/10.1002/ptr.6321>
8. Wang, H., Xia, C., Chen, L., Zhao, J., Tao, W., Zhang, X., Wang, J., Gao, X., Yong, J., & Duan, J. (2019). Phytochemical information and biological activities of quinolizidine alkaloids in *Sophora*: A comprehensive review. *Current Drug Targets*, 20, 1572–1586.
9. Abd-Alla, H. I., Souguir, D., & Radwan, M. O. (2021). Genus *Sophora*: A comprehensive review on secondary chemical metabolites and their biological aspects from past achievements to future perspectives. *Archives of Pharmacal Research*, 44(11), 903–986. <https://doi.org/10.1007/s12272-021-01354-2>
10. Huang, X., Rong, W., Zhang, X., Gao, Y., Zhou, Y., Su, J., Luo, H., Chu, G., & Wang, M. (2024). Transcriptome and metabolome analysis reveal the dynamic changes and biosynthesis pathways of alkaloids in *Sophora alopecuroides* L. under drought stress. *Industrial Crops & Products*, 212, 118365. <https://doi.org/10.1016/j.indcrop.2024.118365>
11. Chacon, D. S., Torres, T. M., da Silva, I. B., de Araújo, T. F., de Araújo Roque, A., Domingos Pinheiro, F. A. S., ... & Cavaleiro, A. J. (2021). *Erythrina velutina* Willd. alkaloids: Piecing biosynthesis together from transcriptome analysis and metabolite profiling of seeds and leaves. *Journal of Advanced Research*, 34, 123–136. <https://doi.org/10.1016/j.jare.2021.01.017>
12. Scotti, M. T., Herrera-Acevedo, C., Oliveira, T., Costa, R., Santos, S., Rodrigues, R., Scotti, L., & Da-Costa, F. (2018). Sistemax, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules*, 23(1), 103.

13. Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (1999). Self-organizing map in Matlab: The SOM Toolbox. In *Proceedings of the Matlab DSP Conference* (pp. 35–40). Espoo, Finland.
14. KNIME. (2024). *KNIME Analytics Platform 5.3.0*. KNIME. <https://www.knime.com/>
15. Majinda, R. R. T. (2018). An update of erythrinan alkaloids and their pharmacological activities. In A. Kinghorn, H. Falk, S. Gibbons, J. Kobayashi, Y. Asakawa, & J. K. Liu (Eds.), *Progress in the chemistry of organic natural products* (Vol. 107). Springer.
https://doi.org/10.1007/978-3-319-93506-5_2
16. Bhargava, V. V., Patel, S. C., & Desai, K. S. (2013). Importance of terpenoids and essential oils in chemotaxonomic approach. *International Journal of Herbal Medicine*, 1(2), 14–21.
17. Al-Tameme, H. J. M. (2015). Chemical profiles as chemotaxonomic tools for some species in Fabaceae in Iraq. *Al-Qadisiyah Journal of Pure Science*, 20(1), 88–99.
18. Park, H.-S., Jayakodi, M., Lee, S. H., Jeon, J.-H., Lee, H.-O., Park, J. Y., Moon, B. C., Kim, C.-K., Wing, R. A., Newmaster, S. G., et al. (2020). Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. *Scientific Reports*, 10, 6112.
<https://doi.org/10.1038/s41598-020-63233-y>
19. Abdelsalam, N. R., Salem, M. Z., Ali, H. M., Mackled, M. I., Mervat, E.-H., Elshikh, M. S., & Hatamleh, A. A. (2019). Morphological, biochemical, molecular, and oil toxicity properties of *Taxodium* trees from different locations. *Industrial Crops and Products*, 139, 111515.
20. Correia, M. V., Scotti, M. T., Ferreira, M. J. P., Alvarenga, S. A. V., Rodrigues, G. V., & Emerenciano, V. P. (2008). Self-organizing maps as a new tool for classification of plants at lower hierarchical levels. *Natural Product Communications*, 3(10).
<https://doi.org/10.1177/1934578X0800301029>
21. Emerenciano, V. P., Barbosa, K. O., Scotti, M. T., & Ferreira, M. J. P. (2007). Self-organizing maps in chemotaxonomic studies of Asteraceae: A classification of tribes using flavonoid data. *Journal of the Brazilian Chemical Society*, 18(5), 801–809.
22. Scotti, M. T., Emerenciano, V., Ferreira, M. J. P., Scotti, L., Stefani, R., da Silva, M. S., & Mendonça Junior, F. J. B. (2012). Self-organizing maps of molecular descriptors for sesquiterpene lactones and their application to the chemotaxonomy of the Asteraceae family. *Molecules*, 17(4), 4684–4702. <https://doi.org/10.3390/molecules17044684>
23. Brown, R. D., & Martin, Y. C. (1998). The information content of 2D and 3D structural descriptors relevant to ligand–receptor binding. *Journal of Chemical Information and Computer Sciences*, 38(6), 983–996.
24. Lewis, G. P., Schrire, B. D., Mackinder, B. A., & Lock, M. (2005). *Legumes of the world*. Royal Botanic Gardens.
25. LPWG (Legume Phylogeny Working Group). (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon*, 66(1), 44–77.
26. Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews*, 63(4), 968–989.

27. Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., & Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*, 13, 162. <https://doi.org/10.1186/1471-2164-13-162>
28. Swain, T. (1963). *Chemical plant taxonomy*. Academic Press.
29. Swain, T. (1966). *Comparative phytochemistry*. Academic Press.
30. Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, 68(22–24), 2831–2846. <https://doi.org/10.1016/j.phytochem.2007.09.017>
31. Bunalema, L., Kirimuhuzya, C., Tabuti, J. R., Waako, P., Magadula, J. J., Otieno, N., ... Okemo, P. (2011). The efficacy of the crude root bark extracts of *Erythrina abyssinica* on rifampicin resistant *Mycobacterium tuberculosis*. *African Health Sciences*, 11(4), 587–593.
32. Devi, R. M., & Manoharan, A. (2011). Characteristics of pharmacognostical significance of *Erythrina variegata* var. and *Ficus racemosa* Linn. bark. *Journal of Chemical and Pharmaceutical Research*, 3(6), 707–714.
33. Rambo, D. F., Biegelmeier, R., Toson, N. S. B., Dresch, R. R., Moreno, P. R. H., & Henriques, A. T. (2019). The genus *Erythrina* L.: A review on its alkaloids, preclinical, and clinical studies. *Phytotherapy Research*, 33(6), 1258–1276. <https://doi.org/10.1002/ptr.6321>
34. Lee, S. T., Cook, D., & Molyneux, R. J. (2014). Identification of the quinolizidine alkaloids in *Sophora leachiana*. *Biochemical Systematics and Ecology*, 54, 1–4. <https://doi.org/10.1016/j.bse.2013.12.020>
35. Wang, L., You, Y., Wang, S., et al. (2012). Synthesis, characterization and in vitro anti-tumor activities of matrine derivatives. *Bioorganic & Medicinal Chemistry Letters*, 22(12), 4100–4102.
36. Wang, H., Xia, C., Chen, L., Zhao, J., Tao, W., Zhang, X., & Duan, J. (2019). Phytochemical information and biological activities of quinolizidine alkaloids in *Sophora*: A comprehensive review. *Current Drug Targets*, 20(6), 1572–1586. <https://doi.org/10.2174/1389450120666190618125816>
37. De Queiroz, K. (1988). Systematics and the Darwinian revolution. *Philosophy of Science*, 55, 238–259.
38. Stevens, P. (1984). Metaphors and typology in the development of botanical systematics 1690–1960, or the art of putting new wine in old bottles. *Taxon*, 33, 169–211.
39. Wink, M., & Schimmer, O. (2010). Molecular modes of action of defensive secondary metabolites. In M. Wink (Ed.), *Functions and biotechnology of plant secondary metabolites* (Annual Plant Reviews, Vol. 39, pp. 21–161). Wiley-Blackwell.
40. Wink, M., & Van Wyk, B.-E. (2008). *Mind-altering and poisonous plants of the world*. Briza Publications.
41. Wink, M., & Mohamed, G. I. A. (2003). Evolution of chemical defense traits in the Leguminosae: Mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the *rbcL* gene. *Biochemical Systematics and Ecology*, 31(8), 897–917.

42. Berkov, S., Osorio, E., Viladomat, F., & Bastida, J. (2020). Chemodiversity, chemotaxonomy and chemoecology of Amaryllidaceae alkaloids. *The Alkaloids: Chemistry and Biology*, 83, 113–174. <https://doi.org/10.1016/bs.alk.2019.10.002>
43. Cavalcanti, E. B. V. S., Scotti, M. T., Scotti, L., & Emerenciano, V. de P. (2017). Application of Self-Organizing Maps generated from Molecular Descriptors of Flavonoid in the Chemotaxonomy of the Asteraceae Family. *MOL2NET, International Conference Series on Multidisciplinary Sciences*, 3. <https://doi.org/10.3390/mol2net-03-XXXX>
44. Bo, W., Chen, L., Qin, D., Geng, S., Li, J., Mei, H., Li, B., & Liang, G. (2021). [Article title not provided]. *Trends in Food Science & Technology*, 114, 176.
45. Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). [Article title not provided]. *QSAR & Combinatorial Science*, 22, 69.
46. Livingstone, D. (2009). *A practical guide to scientific data analysis* (1st ed.). Wiley-VCH.
47. Amaral, F. (2016). *Aprenda mineração de dados* (1ª ed.). Alta Books.
48. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6–7), 476–488. <https://doi.org/10.1002/minf.201000061>
49. Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., & Tropsha, A. (2010). [Article title not provided]. *ACS Nano*, 4(11), 5703.
50. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J. C., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R. D., Benigni, R., Yang, C., Rathman, J. F., Terfloth, L., Gasteiger, J., Richard, A. M., & Tropsha, A. (2014). [Article title not provided]. *Journal of Medicinal Chemistry*, 57, 4977.
51. Menezes, R. P. B. de, Scotti, L., & Scotti, M. T. (2024). Aprendizado de máquina aplicado a QSAR. *Química Nova*, 47(7), e20240024. <https://doi.org/10.21577/0100-4042.20240024>
52. Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812–2831. <https://doi.org/10.1039/c3ay41907j>
53. Martins, J. P. A., & Ferreira, M. M. C. (2013). QSAR modeling: Um novo pacote computacional open source para gerar e validar modelos QSAR. *Química Nova*, 36(4), 589–595. <https://doi.org/10.1590/S0100-40422013000400013>
54. Graham, P. L. (2017). *An introduction to medicinal chemistry* (5th ed.). Oxford University Press.
55. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1–3), 3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
56. Maurer, T. S., Smith, D., Beaumont, K., & Di, L. (2020). Dose selection using physiologically based pharmacokinetic (PBPK) modeling: A review of applications and implications for translational drug development. *Journal of Medicinal Chemistry*, 63(13), 6423–6435. <https://doi.org/10.1021/acs.jmedchem.9b02163>
57. Pereira, D. G. (2007). Importância do metabolismo no planejamento de fármacos. *Química Nova*, 30(1), 259–265. <https://doi.org/10.1590/S0100-40422007000100029>

58. Guido, R. V. C., Andricopulo, A. D., & Oliva, G. (2010). Planejamento de fármacos, biotecnologia e química medicinal: Aplicações em doenças infecciosas. *Estudos Avançados*, 24(70), 181–208. <https://doi.org/10.1590/S0103-40142010000300006>
59. Alves, M. S., Streite, L., & Pizzolato, T. M. (2023). Utilização de modelos in silico para a avaliação da toxicidade de resíduos de agrotóxicos, fármacos e metabólitos em águas naturais. *Química Nova*, 46(9), 881–889. <https://doi.org/10.21577/0100-4042.20230062>
60. World Health Organization. (2021). *Leishmaniasis*. <https://www.who.int/health-topics/leishmaniasis>
61. World Health Organization, et al. (2023). *Global report on neglected tropical diseases 2023*. World Health Organization.
62. World Health Organization. (2022). *Safety in administering medicines for neglected infectious diseases*. Geneva: World Health Organization. <https://apps.who.int/iris/handle/10665/344059>
63. Burrows, J. N., et al. (2014). The role of modern drug discovery in the fight against neglected and tropical diseases. *MedChemComm*, 5(6), 688–700. <https://doi.org/10.1039/c4md00120f>
64. World Health Organization. (2018). *Weekly epidemiological record, no 40, 5 out. 2018*. <http://apps.who.int/iris/bitstream/handle/10665/275333/WER9340.pdf?ua=1>
65. DNDi. (2019). *Rumo a uma nova geração de tratamentos para as leishmanioses*. DNDi. Recuperado de https://dndi.org/wp-content/uploads/2019/09/DNDi_Leishmaniasis_2019_POR.pdf
66. Viana, M. T. B., & Santos, V. M. (2024). Desafios da terapêutica da leishmaniose: Limitações atuais e perspectivas futuras. *Brazilian Journal of Implantology and Health Sciences*, 6(9), 3501–3518. <https://doi.org/10.36557/2674-8169.2024v6n9p3501-3518>
67. Almeida, O. L. S., & Santos, J. B. (2011). Avanços no tratamento da leishmaniose tegumentar do novo mundo nos últimos dez anos: Uma revisão sistemática da literatura. *Anais Brasileiros de Dermatologia*, 86(3), 497–506. <https://doi.org/10.1590/S0365-05962011000300012>
68. Alves, V. M., Braga, R. C., Muratov, E. N., & Andrade, C. H. (2018). Cheminformatics: An introduction. *Química Nova*, 41, 202–212. <https://doi.org/10.21577/0100-4042.20180042>
69. Chemaxon. (2023). *Marvinsketch* (v. 23.14) [Software]. Chemaxon. <https://chemaxon.com/marvin>
70. Wavefunction, Inc. (2014). *Spartan* (Versão 14) [Software]. Wavefunction, Inc.
71. Dufe, V. T., Ingner, D., Heby, O., Khomutov, A. R., Persson, L., & Al-Karadaghi, S. (2007). A structural insight into the inhibition of human and *Leishmania donovani* ornithine decarboxylases by 1-amino-oxy-3-aminopropane. *Biochemical Journal*, 405(2), 261–268. <https://doi.org/10.1042/BJ20070423>
72. Gabison, L., Chiadmi, M., El Hajji, M., Castro, B., Colloc'h, N., & Prange, T. (2010). Near-atomic resolution structures of urate oxidase complexed with its substrate and analogues: The protonation state of the ligand. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), 714–724. <https://doi.org/10.1107/S0907444910005136>

73. Gazanion, E., Garcia, D., Silvestre, R., Gérard, C., Guichou, J. F., Labesse, G., ... & Vergnes, B. (2011). The *Leishmania* nicotinamidase is essential for NAD⁺ production and parasite proliferation. *Molecular Microbiology*, 81(5), 1165–1178. <https://doi.org/10.1111/j.1365-2958.2011.07717.x>
74. Are, S., Gatreddi, S., Jakkula, P., & Qureshi, I. A. (2020). Structural attributes and substrate specificity of pyridoxal kinase from *Leishmania donovani*. *International Journal of Biological Macromolecules*, 150, 1240–1248. <https://doi.org/10.1016/j.ijbiomac.2020.02.102>
75. Molexus. (2013). *Molegro Virtual Docker* (Versão 6.0) [Software]. Molexus.
76. Biovia. (2023). *Discovery Virtual Studio* [Software]. Biovia.
77. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480. <https://doi.org/10.1109/5.58325>
78. Silva, M. L. M. (2017). *Uma introdução aos mapas neurais auto-organizáveis: As redes de Kohonen* (Trabalho de conclusão de curso). Universidade Católica de Brasília, Brasília, DF, Brasil. Disponível em: <https://repositorio.ucb.br:9443/jspui/bitstream/123456789/12554/1/MariaLuizaMoraisdaSilvaTCCGraduacao2017.pdf>
79. Silva, H. C. (2014). *Aplicação de redes neurais artificiais do tipo mapas auto-organizáveis: Estudo de casos* (Dissertação de mestrado). Universidade Estadual de Londrina, Londrina, Paraná, Brasil.
80. Guha, R., Serra, J. R., & Jurs, P. C. (2004). Generation of QSAR sets with a self-organizing map. *Journal of Molecular Graphics and Modelling*, 23(1), 1–14. <https://doi.org/10.1016/j.jmgm.2004.01.001>
81. Correia, M. V., Scotti, M. T., Ferreira, M. J. P., Alvarenga, S. A. V., Rodrigues, G. V., & Emerenciano, V. P. (2008). Self-organizing maps as a new tool for classification of plants at lower hierarchical levels. *Natural Product Communications*, 3(10), 1723–1730.
82. Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 24(2), 195. <https://doi.org/10.1038/nbt1386>
83. Drew, P. J., & Monson, J. R. T. (2000). Artificial neural networks. *Surgery*, 127(1), 3–11. <https://doi.org/10.1067/msy.2000.102953>
84. Abdolmaleki, A., Shiri, F., & Ghasemi, J. B. (2021). Use of molecular docking as a decision-making tool in drug discovery. In *Molecular docking for computer-aided drug design* (pp. 229-243). Academic Press. <https://doi.org/10.1016/B978-0-12-819730-5.00014-3>
85. Lyne, P. D. (2002). Structure-based virtual screening: An overview. *Drug Discovery Today*, 7(20), 1047–1055. [https://doi.org/10.1016/S1359-6446\(02\)02274-3](https://doi.org/10.1016/S1359-6446(02)02274-3)
86. Du, X., et al. (2016). Insights into protein–ligand interactions: Mechanisms, models, and methods. *International Journal of Molecular Sciences*, 17(2), 144. <https://doi.org/10.3390/ijms17020144>
87. Cheng, T., et al. (2012). Structure-based virtual screening for drug discovery: A problem-centric review. *The AAPS Journal*, 14, 133–141. <https://doi.org/10.1208/s12248-012-9367-3>
88. Ibrahim, T. M., Bauer, M. R., & Boeckler, F. M. (2015). Applying DEKOIS 2.0 in structure-based virtual screening to probe the impact of preparation procedures and score

normalization. *Journal of Cheminformatics*, 7, 1–16. <https://doi.org/10.1186/s13321-015-0084-1>

89. Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7(1), 42717. <http://www.swissadme.ch/>

90. Tran, H. T., Nguyen, T. T., & Venkatesh, S. (2021). Deep-PK: A web-based tool for predicting pharmacokinetics of small molecules using deep learning. *Bioinformatics Laboratory, University of Queensland*. <https://biosig.lab.uq.edu.au/deeppk/>

91. ChemAxon. (2025). *Standardizer*. ChemAxon. Disponível em https://docs.chemaxon.com/display/docs/standardizer_installation-and-system-requirements.md

92. Brooks, H. B., & Phillips, M. A. (1997). Characterization of the reaction mechanism for *Trypanosoma brucei* ornithine decarboxylase by multiwavelength stopped-flow spectroscopy. *Biochemistry*, 36(48), 15147–15155. <https://doi.org/10.1021/bi9718108>

93. Arnott, J. A., & Planey, S. L. (2012). The influence of lipophilicity in drug discovery and design. *Expert Opinion on Drug Discovery*, 7(10), 863–875. <https://doi.org/10.1517/17460441.2012.721715>

94. Waring, M. J. (2010). Lipophilicity in drug discovery. *Expert Opinion on Drug Discovery*, 5(3), 235–248. <https://doi.org/10.1517/17460441.2010.485506>

95. Kibet, S., et al. (2024). Unveiling the potential of Ent-Kaurane diterpenoids: Multifaceted natural products for drug discovery. *Pharmaceuticals*, 17(4), 510. <https://doi.org/10.3390/ph17040510>

96. Ahire, E. D., et al. (2021). Drug discovery, drug-likeness screening, and bioavailability: Development of drug-likeness rule for natural products. In *Applied pharmaceutical practice and nutraceuticals* (pp. 191-208). Apple Academic Press.

97. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 45(12), 2615–2623. <https://doi.org/10.1021/jm020017n>

98. Chen-Yang, J., Jia, J.-Y., Li, G.-F., Hao, G.-F., & Yang, G. (2020). A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drug Discovery Today*, 25(1), 248–258. <https://doi.org/10.1016/j.drudis.2019.10.014>

99. Ghose, A. K., Vellarkad, N. V., & John, J. W. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, 1(1), 55–68. <https://doi.org/10.1021/cc980024n>

100. Muegge, I., Heald, S. L., & Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, 44(12), 1841–1846. <https://doi.org/10.1021/jm010431e>

101. Ursu, O., Rayan, A., Goldblum, A., & Oprea, T. I. (2011). Understanding drug-likeness. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(1), 760–781. <https://doi.org/10.1002/wcms.1469>

102. Polinsky, A. (2008). Lead-likeness and drug-likeness. In *The practice of medicinal chemistry* (pp. 244–254). Academic Press.

103. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). *Adv. Drug Delivery Reviews*, 46, 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00130-3](https://doi.org/10.1016/S0169-409X(00)00130-3)
104. Are, S., Gatreddi, S., Jakkula, P., & Qureshi, I. A. (2020). Structural attributes and substrate specificity of pyridoxal kinase from *Leishmania donovani*. *International Journal of Biological Macromolecules*, 152, 812–827. <https://doi.org/10.1016/j.ijbiomac.2020.02.138>
105. Kohonen, T. (2012). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>
106. Ramos, Y. J., Brito-Machado, D., Queiroz, G. A., Guimarães, E. F., Defaveri, A. C. A., & Moreira, D. (2020). Chemical composition of the essential oils of circadian rhythm and of different vegetative parts from *Piper mollicomum* Kunth — a medicinal plant from Brazil. *Biochemical Systematics and Ecology*, 92, 104–116. <https://doi.org/10.1016/j.bse.2020.104116>
107. Ramos, Y. J., Gouvêa-Silva, J. G., Machado, D. B., Felisberto, J. S., Pereira, R. C., Sadgrove, N. J., & Moreira, D. L. (2023). Chemophenetic and chemodiversity approaches: New insights on modern study of plant secondary metabolite diversity at different spatiotemporal and organizational scales. *Revista Brasileira de Farmacognosia*, 33(3), 49–72. <https://doi.org/10.1007/s43450-022-00327-w>
108. J, B., BM, M., & Chanda, K. (2021). An overview on the therapeutics of neglected infectious diseases - Leishmaniasis and Chagas diseases. *Frontiers in Chemistry*, 2021. <https://doi.org/10.3389/fchem.2021.728469>
109. Silva-López, R. E. da. (2010). Proteases de *Leishmania*: novos alvos para o desenvolvimento racional de fármacos. *Química Nova*, 33(7), 1590–1596. <https://doi.org/10.1590/S0100-40422010000700022>.
110. Tabrez, S., Rahman, F., Ali, R., Akand, S. K., Alaidarous, M. A., Alshehri, B. M., Banawas, S., Bin Dukhayil, A. A., & Rub, A. (2021). Targeting sterol alpha-14 demethylase of *Leishmania donovani* to fight against leishmaniasis. *Journal of Cellular Biochemistry*, Advance online publication. <https://doi.org/10.1002/jcb.29922>
111. Gazanion, E., Garcia, D., Silvestre, R., Gérard, C., Guichou, J. F., Labesse, G., Seveno, M., Cordeiro-Da-Silva, A., Ouaisi, A., Sereno, D., & Vergnes, B. (2011). The *Leishmania* nicotinamidase is essential for NAD⁺ production and parasite proliferation. *Cellular Microbiology*, 13(10), 1418–1434. <https://doi.org/10.1111/j.1365-2958.2011.07799.x>
112. Are, S., Gatreddi, S., Jakkula, P., & Qureshi, I. A. (2020). Structural attributes and substrate specificity of pyridoxal kinase from *Leishmania donovani*. *International Journal of Biological Macromolecules*, 152, 812–827. <https://doi.org/10.1016/j.ijbiomac.2020.02.257>

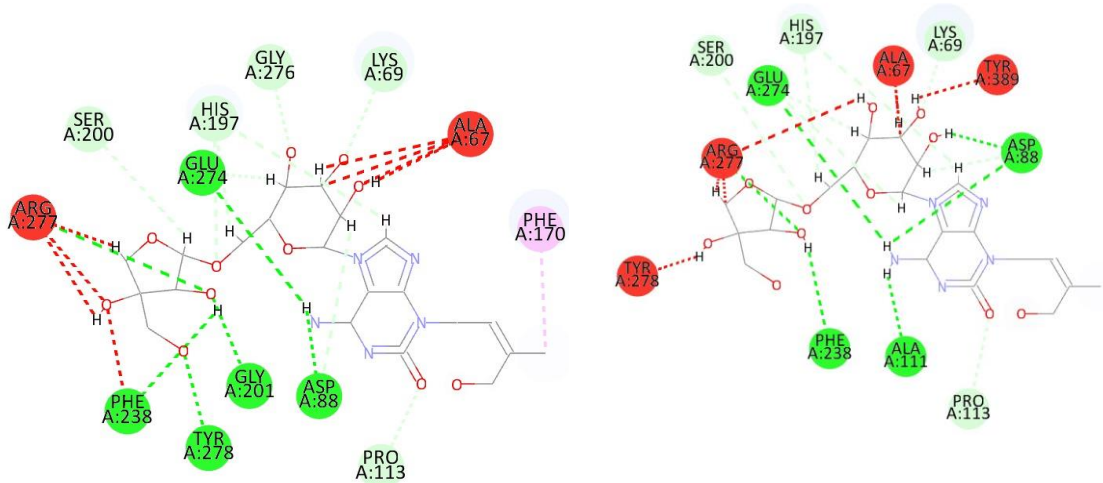
APÊNDICES

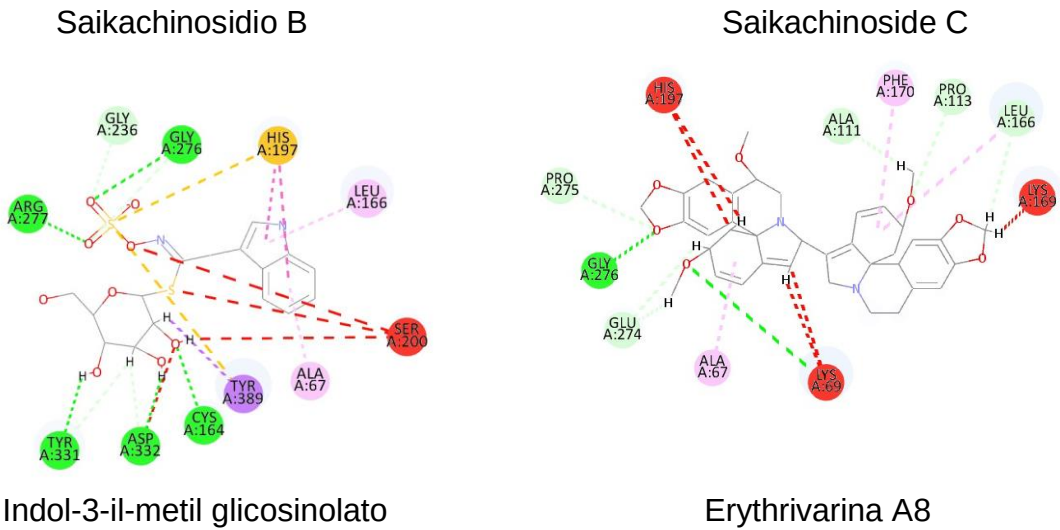
APÊNDICES

APÊNDICE I

Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
Ornitina decarboxilase	2O00	Pentamidina	-103,92	0.14 Å
		Erythrivarina A5	-151,03	
		Saikachinosidio B	-145,68	
		Indol-3-il-metil glicosinolato	-144.83	
		Saikachinoside C	-144,73	
		Erythrivarina A8	-135,46	
		Erythrivarina A1	-134.79	
		Erythrivarina C	-134.48	
		Prosojuliflavona	-134.16	
		Yuremamina	-134.00	
		Erythritol	-133.88	

APÊNDICE II

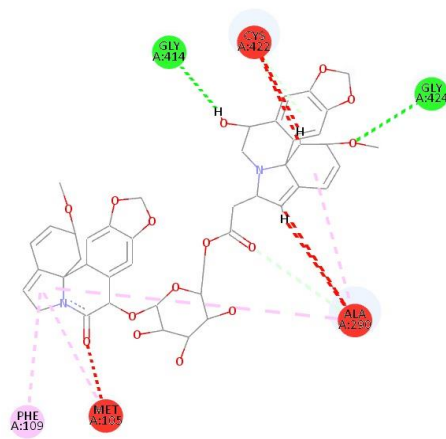




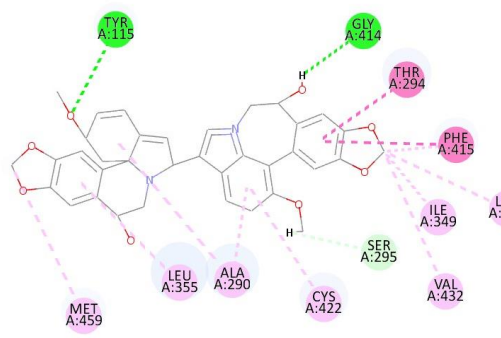
APÊNDICE III

Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
Urato Oxidase	3LD4	Pentamidina	-83,51	0.22 Å
		Eritrivarina D	-194,60	
		Eritrivarina C	-190,96	
		Eritrivarina B	-163.10	
		Eritrivarina U	-156,93	
		Eritrivarina F	-156,35	
		Eritrivarina E	-152.89	
		Dissulfato de Locustosidio B	-152.20	
		Eritrivarina Q	-144.47	
		Eritrivarina R	-143.57	
		Eritrivarina A12	-141.54	

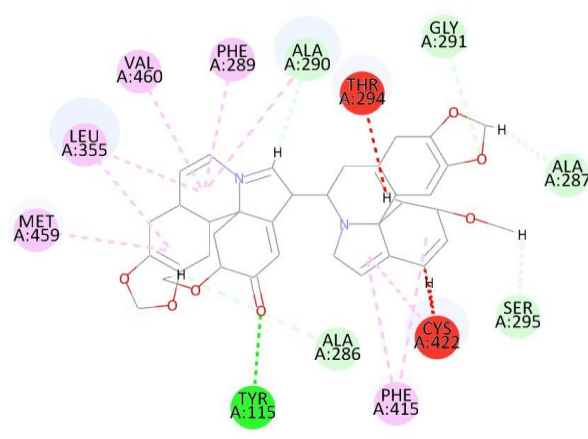
APÊNDICE IV



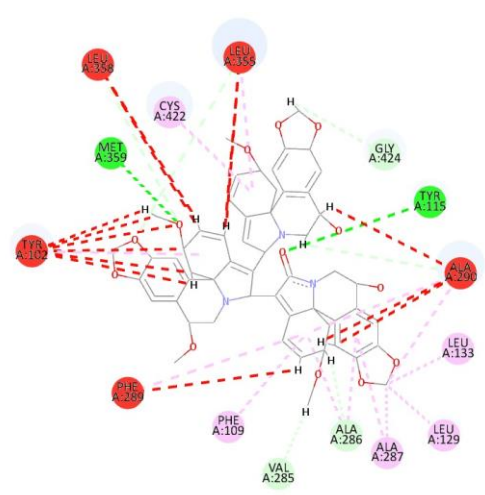
Eritrivarina C



Eritrivarina B



Eritrivarina U



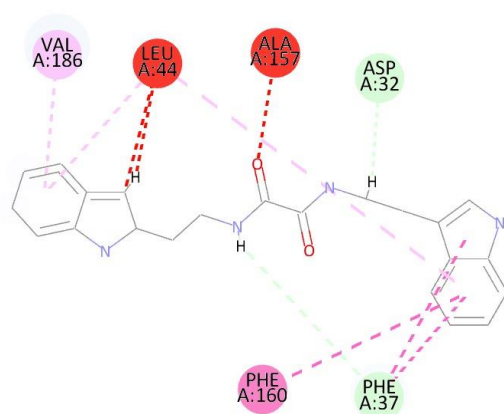
Eritrivarina F

APÊNDICE V

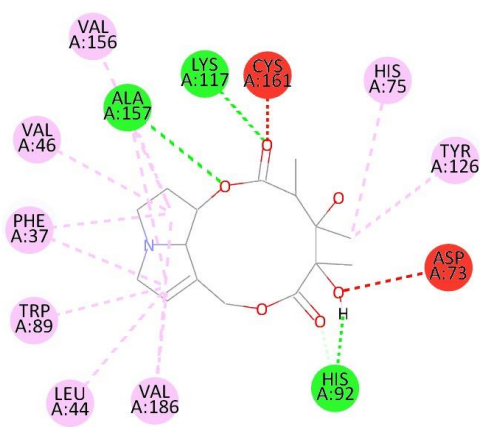
Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
PnC1	3R2J	Pentamidina	-78,36	0.28 Å
		Vicina	-124,46	
		Monocrotalina	-117,69	
		Erisopina	-113,76	

N,N-Bis(2-(1 <i>H</i> -indol-3-il) etil) oxalamida	-113,51
Convicina	-110,47
(+)-(trans-4'-hidroxi-3'-metoxicinnamoil) epilupinina	-107.27
Convicina I	-105.33
12α-Hidroxisophocarpina	-105.11
Bufotenina	-104.21
Calpurnina	-102.00

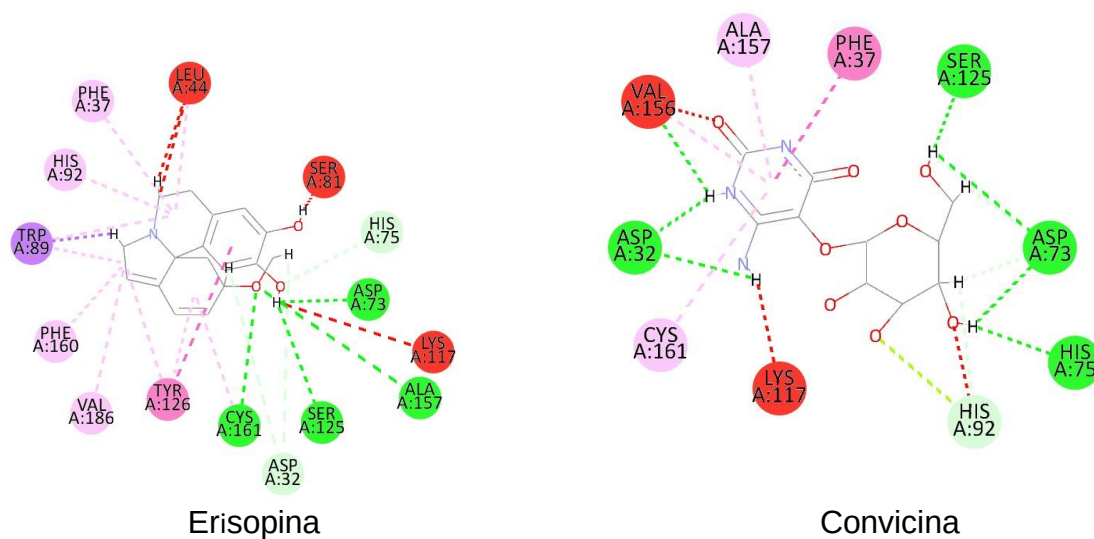
APÊNDICE VI



N,N-Bis(2-(1*H*-indol-3-il) etil) oxalamida



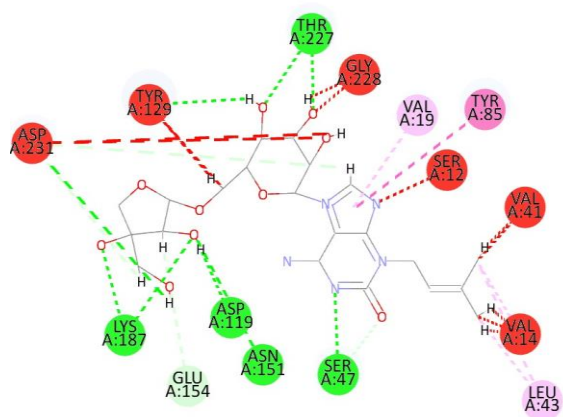
Monocrotalina



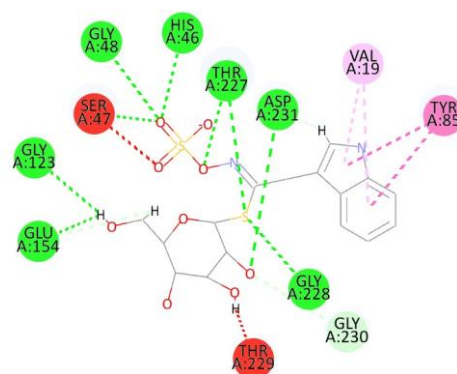
APÊNDICE VII

Alvo	PDB ID	Ligante ID	MolDock Score	RMSD
Piridoxal quinase	6K91	Pentamidina	-85,18	0.10
		Saikachinosidio B	-174,61	
		Locustosidio B	-170,82	
		Prosojuli	-166,40	
		Indol-3-il-metil glicosinolato	-162,92	
		Juliprosinina	-162,63	
		Eritrivarina C	-162,58	
		(+)-3-O-feruloilcassino	-157,59	
		Saikachinosidio C	-157,52	
		Dissulfato de Locustosidio B	-154,58	
		Eritrivarina A13	-152,01	

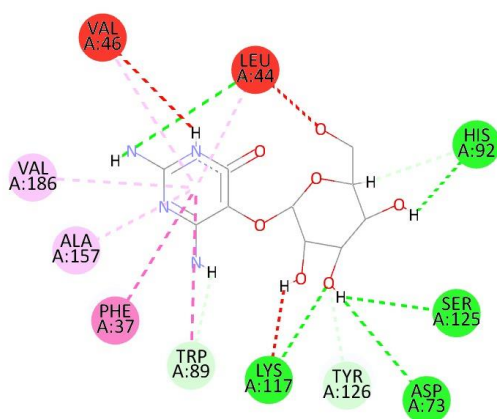
APÊNDICE VIII



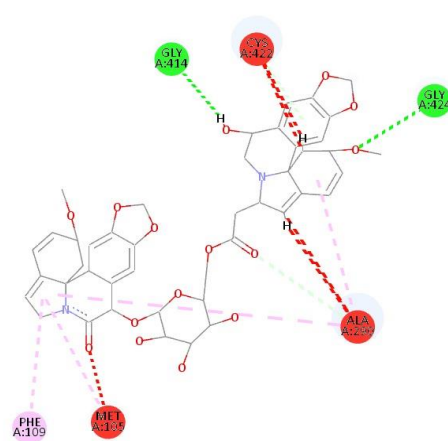
Locustosidio B



Prosojulinina



Indol-3-il-metil glicosinolato



Juliprosinina

Catálogo na publicação
Seção de Catalogação e Classificação

P814e Pontes, Adiel Henrique de Oliveira.

Estudos quimioinformáticos de alcaloides de fabaceae
: quimiotaxonomia, atividades biológicas e
antileishmania / Adiel Henrique de Oliveira Pontes. -
João Pessoa, 2025.
90 f. : il.

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

1. Alcaloides. 2. Quimiotaxonomia. 3. Metabolitos
secundários. I. Scotti, Marcus Tullius. II. Título.

UFPB/BC

CDU 547.94(043)