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

ANDREZA BARBOSA SILVA CAVALCANTI

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE DITERPENOS ISOLADOS DA FAMÍLIA LAMIACEAE COM POTENCIAL ATIVIDADE ANTITUBERCULAR

João Pessoa-PB
2021

ANDREZA BARBOSA SILVA CAVALCANTI

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE DITERPENOS ISOLADOS DA FAMÍLIA LAMIACEAE COM POTENCIAL ATIVIDADE ANTITUBERCULAR

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

ORIENTADOR: Prof. Dr. Marcus Tullius Scotti

COORIENTADOR: Prof. Dr. Vicente Carlos de Oliveira Costa

João Pessoa-PB
2021

Catalogação na publicação
Seção de Catalogação e Classificação

C376e Cavalcanti, Andreza Barbosa Silva.
Estudos quimiotaxonômicos e triagem virtual de diterpenos isolados da família Lamiaceae com potencial atividade antitubercular / Andreza Barbosa Silva Cavalcanti. - João Pessoa, 2021.
175 f. : il.

Orientação: Marcus Tullius Scotti.
Coorientação: Vicente Carlos de Oliveira Costa.
Tese (Doutorado) - UFPB/CCS.

1. Lamiaceae - Quimiotaxonomia. 2. Hyptidinae. 3. Leptohyptis macrostachys. 4. Mesosphaerum sidifolium. 5. Mycobacterium tuberculosis. 6. Modelos de predição. I. Scotti, Marcus Tullius. II. Costa, Vicente Carlos de Oliveira. III. Título.

UFPB/BC CDU 582.929.4(043)

Elaborado por WALQUELINE DA SILVA ARAUJO - CRB-15/514

ANDREZA BARBOSA SILVA CAVALCANTI

ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE DITERPENOS ISOLADOS DA FAMÍLIA LAMIACEAE COM POTENCIAL ATIVIDADE ANTITUBERCULAR

COMISSÃO EXAMINADORA

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

Prof. Dr. Massuo Jorge Kato
PhD em Química Orgânica
Universidade de São Paulo
Avaliador externo

Prof. Dr. Sócrates Golzio dos Santos
PhD em Produtos Naturais e Sintéticos Bioativos
Universidade Federal da Paraíba
Avaliador interno

Prof. Dr. Francisco Jaime B. Mendonça Junior
PhD em Ciências Biológicas
Universidade Federal de Pernambuco
Avaliador interno

Prof. Dr. Wallace Duarte Fragoso
PhD em Química
Universidade Federal de Pernambuco
Avaliador externo

Dedicatória

Aos meus pais, Luzinete Barbosa e Reginaldo Silva *(In memorian)* por todo cuidado e amor.

Ao meu esposo, Edilson Cavalcanti pelo amor e companheirismo.

Ao meu filho, Bernardo por esse amor infinito.

DEDICO

Agradecimentos

À **Deus**, por tudo que Ele concretiza diariamente em minha vida, por me permitir saúde para que eu consiga buscar sonhos e realizações.

A Nossa Senhora das Graças, minha fiel intercessora, pelo seu exemplo de fé e coragem.

Aos meus pais, Luzinete e Reginaldo (*In memorian*) pelo exemplo que sempre foram pra mim, por todo incentivo para estudar e conseguir prosseguir nas etapas da vida, vocês são os melhores pais que eu poderia ter.

Ao meu esposo Edilson, por todo apoio, compreensão, carinho, amor e por viver comigo e com nosso pequeno a plenitude da palavra chamada família, você me faz muito feliz.

Ao meu filho, Bernardo pelo amor incondicional, infinito, combustível diario.

Ao meu irmão Júnior, por toda colaboração, paciência, carinho e respeito a mim concedidos.

A toda minha família e amigos pelas orações, incentivo e torcida.

Ao meu orientador, Prof. Dr. Marcus Tulius Scotti, por todo apoio que me concedeu durante uma trajetória tão importante na minha vida, por ensinar e dar suporte sempre que necessário como também pela oportunidade de desenvolver este trabalho.

Ao meu coorientador Prof. Dr. Vicente Carlos de Oliveira Costa, por todo apoio, paciência, ensinamentos e suporte.

Aos Professores Dr. Marcelo Sobral da Silva e Dr. Josean Fechine Tavares, que abriram as portas do caminho acadêmico para mim, sempre incentivando e dando suporte.

A Profa. Dra. Luciana Scotti por todo apoio e incentivo, pelo exemplo de pessoa que representa.

Aos membros da banca avaliadora, por aceitarem o convite e contribuir com o meu trabalho.

À Profa Dra. Maria de Fátima Agra pela identificação do material vegetal em estudo.

Aos meus amigos de laboratório: Renata, Gabriela, Natália, Chonny, Luana, Mayara, Pedro, Edleuza, Ana Rita, Jociano.

Aos meus amigos da turma de doutorado: Rose, Laiane, Anderson, Kaio, Lucas, Yuri, Márcio, Diego e Denise por todos os momentos compartilhados, e pelo apoio sempre.

A todos os companheiros de bancada, alunos de iniciação científica e pós-graduandos que fazem parte do Laboratório de Fitoquímica (Sobral e Fechine), pela amizade.

A técnicos de laboratório Nonato, Evandro e Marcelo, pela contribuição imprescindível para a conclusão deste trabalho.

Aos professores do Programa de Pós-graduação pelos ensinamentos transmitidos, em especial a Prof. Dra. Mª de Fátima Vanderley por toda a contribuição de ensinamentos e incentivos desde a graduação.

À Juliana, Rose, Sabrina, Laiane e Renata, pela amizade, por todos os conselhos e contribuições.

As minhas amigas Gilcélia, Antônia, Luara e Georgia por todo apoio, incentivo e confiança.

À família Cavalcanti pelo incentivo, carinho e alegrias.

À Universidade Federal da Paraíba.

Ao Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Enfim, a todos que colaboraram de alguma forma para a concretização desta etapa na minha vida.

GRATIDÃO!!!!!

Andreza Barbosa Silva Cavalcanti

O que eu faço, é uma gota no meio de um oceano. Mas sem ela, o oceano será menor.

(Madre Tereza de Calcutá)

Resumo

CAVALCANTI, A. B. S.[1] **ESTUDOS QUIMIOTAXONÔMICOS E TRIAGEM VIRTUAL DE DITERPENOS ISOLADOS DA FAMÍLIA LAMIACEAE COM POTENCIAL ATIVIDADE ANTITUBERCULAR**. Tese (Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos) – Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, p.175, 2021.

Resumo

A família Lamiaceae apresenta distribuição cosmopolita, sendo composta por uma grande variedade de metabólitos secundários, possui mais de 290 gêneros 7000 espécies. Estima-se que no Brasil essa família esteja representada por 36 gêneros e 490 espécies. Diante da grande diversidade da família Lamiaceae, o objetivo deste trabalho foi construir um banco de dados de diterpenos, realizar estudos quimiotaxonômicos auxiliando na prospecção de diterpenos com características estruturais específicas, como também através da triagem virtual, selecionando os mais promissores, e por fim validar por meio do teste atividade biológica. No capítulo 2 foi construído um banco de dados de diterpenos de Lamiaceae e com ele foi realizada uma análise quimiotaxonômica entre as subfamílias de Lamiaceae, utilizando descritores moleculares e mapas auto organizáveis (SOMs). O levantamento de dados evidenciou 4.115 diterpenos diferentes correspondendo a 6.386 ocorrências botânicas, que estão distribuídas em oito subfamílias, 66 gêneros, 639 espécies diferentes e 4880 localizações geográficas, foram adicionados à SistematX. Em todos os mapas obtidos, observou-se uma taxa de acerto superior a 80%, demonstrando uma separação das subfamílias de Lamiaceae, corroborando com os dados morfológicos e moleculares propostos por Li et al. Portanto, por meio deste estudo quimiotaxonômico, foi possível predizer a localização de um diterpeno em uma subfamília e auxiliar na busca de metabólitos secundários com características estruturais específicas, como compostos com potencial atividade biológica. O capítulo 3 apresenta uma avaliação *in silico* do diterpeno labdano ácido labda-8 (17), 14-dien-9,13-epoxi-12β-hidroxi-19-oico, denominado hyptenol (1), juntamente com cinco outros compostos: eritroxilol B (2), acetato de ácido oleanólico (3), ácido betulínico (4), ácido tormêntico (5) e ácido *ent*-3β-acetoxi-kaur-15-en-17-óico (6) que são constituintes químicos das partes aéreas de *Leptohyptis macrostachys*. Foi criado um modelo preditivo para determinar a suscetibilidade à microbactéria da tuberculose, com avaliações moleculares e, em seguida, realizou-se análises *in vitro*. Os resultados demonstraram que o ácido *ent*-3β-acetoxi-kaur-15-en-17-óico (6) apresenta atividade biológica significativa frente ao *Mycobacterium tuberculosis*. No capítulo 4, foi realizado a análise fitoquímica da espécie *Mesosphaerum sidifolium*, pertencente a subtribo Hyptidinae, como também uma triagem virtual do banco de dados de diterpenos dessa tribo, selecionou-se então os compostos mais promissores com atividade antiturbecular frente a *M. tuberculosis*. O estudo *in silico* evidenciou que dos 68 diterpenos de Hyptidinae analisados, 48 apresentaram probabilidade de apresentarem atividade biológica contra *M. tuberculosis* e, por meio do teste biológico, verificou-se que os diterpenos podem ser avaliados posteriormente como moléculas potenciais antituberculares.

Palavras-chave: Lamiaceae, Hyptidinae, *Leptohyptis macrostachys; Mesosphaerum sidifolium*, *Mycobacterium tuberculosis*, Quimiotaxonomia, Triagem virtual, Modelos de predição.

Abstract

CAVALCANTI, A. B. S.[1] **CHEMOTAXONOMIC STUDIES AND VIRTUAL SCREENING OF DITERPENES ISOLATED FROM THE FAMILY LAMIACEAE WITH POTENTIAL ANTITUBERCULAR ACTIVITY.** Thesis - (Postgraduate Program in Natural and Bioactive Synthetic Products) - Health Sciences Center, Federal University of Paraíba, João Pessoa, 175 p.2021.

ABSTRACT

The Lamiaceae family has a cosmopolitan distribution, being composed of a wide variety of secondary metabolites, has more than 290 genera and 7000 species. It is estimated that in Brazil this family is searched for 36 genera and 490 species. Given the great diversity of the Lamiaceae family, the objective of this work was to create a database of diterpenes, carry out auxiliary chemotaxonomic studies in the prospection of diterpenes with specific structural characteristics, as well as virtual screening, selecting the most promising, and finally validate through of the biological activity test. In chapter 2, a database of diterpenes from Lamiaceae was built and with it a chemotaxonomic analysis was performed among the subfamilies of Lamiaceae, using molecular descriptors and self-organizing maps (SOMs). The data collection showed 4115 different diterpenes corresponding to 6386 botanical occurrences, which are distributed in eight subfamilies, 66 genera, 639 different species and 4880 geographic locations were added to SistematX. In all maps obtained, a hit rate of more than 80% was observed, demonstrating a separation of the subfamilies of Lamiaceae, corroborating the morphological and molecular data proposed by Li et al. Therefore, through this chemotaxonomic study, it was possible to predict the location of a diterpene in a subfamily and assist in the search for secondary metabolites with specific structural characteristics, such as compounds with potential biological activity. Chapter 3 presents an *in silico* evaluation of the labdane diterpene acid, labda-8 (17), 14-dien-9,13-epoxy-12β-hydroxy-19-oic acid, called hyptenol (1), together with five other compounds: erythroxylol B (2), oleanolic acid acetate (3), betulinic acid (4), tormentic acid (5) and *ent*-3β-acetoxy-kaur-15-en-17-oic acid (6) which are chemical constituents of the aerial parts of *Leptohyptis macrostachys*. A predictive model was created to determine the susceptibility to the tuberculosis microbacteria, with molecular evaluations, and then *in vitro* analyzes were performed of secondary metabolites with specific structural characteristics, such as compounds with potential biological activity. The results showed that *ent*-3β-acetoxy-kaur-15-en-17-oic acid (6) has significant biological activity against *Mycobacterium tuberculosis*. In chapter 4, the phytochemical analysis of the species *Mesosphaerum sidifolium*, belonging to the subtribe Hyptidinae, was performed, as well as a virtual screening of the diterpene database of this tribe, selecting the most promising compounds with antitubercular activity against *M. tuberculosis*. The *in silico* study showed that of the 68 diterpenes of Hyptidinae analyzed, 48 were likely to have biological activity against *M. tuberculosis* and, through the biological test, it was found that the diterpenes can be evaluated later as potential antitubercular molecules.

Keywords: Lamiaceae, Hyptidinae, *Leptohyptis macrostachys*; *Mesosphaerum sidifolium*, *Mycobacterium tuberculosis*, Chemotaxonomy, Virtual screening, Prediction models.

LISTA DE FIGURAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3

LISTA DE TABELAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3

CAPÍTULO 4

LISTA DE ABREVIATURAS, FÓRMULAS, SIGLAS E SÍMBOLOS

AcOEt	Acetato de etila
ADMET	Absorção, distribuição, metabolismo, excreção e toxicidade
Aju	Ajugoideae
Ala	Alanina
ANNs	Artificial Neural Networks
APT	“Attached Proton Test”
Arg	Arginina
Asn	Asparagina
Asp	Ácido aspártico
AUC	Area under the curve (Área sob a curva)
CADD	Computer Aided Drug Design (Projeto de drogas auxiliado por computador)
Cal	Callicarpoideae
CC	Cromatografia em Coluna
CCDA	Cromatografia em Camada Delgada Analítica
CCEN	Centro de Ciências Exatas e da Natureza
$CDCl_3$	Clorofórmio deuterado
CE_{50}	Concentração da metade máxima efetiva
ChEBI	Chemical Entities of Biological Interest (Laboratório Europeu de Biologia Molecular)
CO2	Dióxido de carbono
COSY	“Correlation Spectroscopy”

Cys	Cisteína
d	Dubleto
dd	Duplo dubleto
ddd	Duplo duplo dubleto
dddd	Duplo duplo duplo dubleto
DNP	Dictionary of Natural Products
EEB	Extrato etanólico bruto
EtOH	Etanol
g	Grama
h	Hora
HMBC	“Heteronuclear Multiple Bond Correlation”
HMQC	“Heteronuclear Multiple Quantum Coherence”
IV	Infravermellho
J	Constante de acoplamento
KBr	Brometo de potássio
kg	Quilograma
k-NN	algoritmo de vizinhos mais próximos
Lam	Lamioideae
LM-EtOH$_{PA}$	Extrato etanólico bruto obtido das partes aéreas de *L. macrostachys*
m	Multipleto
Marinlit	Natural marine Products
MeOD	Metanol deuterado
MeOH	Metanol

mg	Miligrama
MHz	Megahertz
nArCCOR	Number of aromatic esters
nArOH	Number of aromatic hydroxyls
Nep	Nepetoideae
nFuranes	number of furans
NN	redes neurais
NOESY	“Nuclear Overhauser Enhancement Spectroscopy”
nR = Cp	number of primary C terminals - sp2
NRS	Number of ring systems
NuBBE	Bioassay Nucleus, Biosynthesis and Ecophysiology of Natural Products
O-056	(Alcohol) atom-centered fragments
p.	Página
PB	Paraíba
Per	Peronematoideae
PPgPNSB	Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos
ppm	Partes por milhão
Pre	Premnoideae
q	Quadrupleto
RF	*Randon Forest*
Rfs	Fatores de retenção
RMN	Ressonância Magnética Nuclear
RMN ^{13}C	Ressonância Magnética Nuclear de carbono treze

RMN ^{1}H	Ressonância Magnética Nuclear de hidrogênio
s	Singleto
Scu	Scutellarioideae
sl	Singleto largo
SOM	Self-Organizing Map
SVM	*support vector machine*
t	Tripleto
tl	Tripleto largo
UFPB	Universidade Federal da Paraíba
Vit	Viticoideae
δ	Deslocamento químico

SUMÁRIO

Capítulo 1

1 INTRODUÇÃO

Historicamente, as plantas têm sido documentadas por seus usos medicinais, como fonte de tratamento, cura e prevenção de doenças. Elas têm evoluído e se adaptado ao longo do tempo, sendo assim, vem produzindo metabólitos secundários estruturalmente diversos para atuar como mecanismo de defesa de predadores (CASANOVA; COSTA, 2017; DIAS; URBAN; ROESSNER, 2012).

Atualmente são utilizadas a associação de várias áreas estratégicas em ciência e tecnologia para o desenvolvimento de novas substâncias candidatas a fármacos por ser um processo complexo e de alto custo (AMARAL et al., 2017). Mas a química de produtos naturais representa historicamente uma destas ferramentas de sucesso (BORGES et al., 2017). Sendo assim, os pesquisadores vêm buscando descobrir novas drogas oriundas das plantas, através da validação do uso etnofarmacológico de metabólitos secundários (MAMADALIEVA et al., 2021).

O Brasil por sua vez, está situado em uma posição singular para a ciência de produtos naturais, em decorrência de apresentar a maior biodiversidade do mundo, com mais de 55.000 espécies de plantas, que compõe uma riqueza de substâncias com grande potencial terapêutico (BORGES et al., 2017; SÁ- FILHO et al., 2021). Sendo o bioma caatinga exclusivamente brasileiro, ocorrendo quase predominantemente na região nordeste, portando uma vegetação única, conhecida como uma verdadeira farmácia viva na qual suas espécies tem sido utilizada pela população na medicina tradicional (DRUMOND; SCHISTEK; SEIFFARTH, 2012; VIEIRA DE BARROS et al., 2021). Diante disso é crescente o interesse da indústria farmacêutica pelas pesquisas com produtos naturais obtidos dessas plantas, visto que mesmo diante de uma grande diversidade natural o Brasil é provedor de poucos produtos ativos considerados inovadores (CALIXTO, 2019).

Na busca desses metabólitos secundários, podemos utilizar ferramentas de desreplicação, que é o acesso a características das moléculas já relatadas na literatura e que estão disponíveis em bancos de dados virtuais. Então, essas bases de dados podem fornecer diversas informações a respeito dos compostos, como dados biológicos, biogeográficos e taxonômicos, ou a presença de um certo composto (novo ou conhecido) em outros indivíduos da mesma espécie, gênero, subfamília e família.

Atualmente temos disponível uma interface web que é o SISTEMATX web *(*https://sistematx.ufpb.br*)*, que disponibiliza um banco de dados de metabólitos secundários, apresentando uma riqueza de informações úteis para a comunidade científica sobre os produtos naturais, dentre elas está a pesquisa por estrutura química, código SMILES, nome do composto como também informações específicas para classificação taxonômica (de família para espécie) da planta da qual o composto foi isolado e a localização de espécies (SCOTTI et al., 2018).

Esse é um exemplo que a quimioinformática vem evoluindo de modo interdisciplinar com aplicações ilimitadas na exploração de bases de dados químicos e descoberta de novos constituintes. Resultando em modelos computacionais que são usados para predizer a atividade de moléculas que precisam de dados experimentais, ou seja, dados de ensaios *in vitro* e *in vivo* (ALVES et al., 2017). Portanto a utilização de técnicas analíticas associados a métodos computacionais pode fornecer o isolamento e identificação de novas moléculas bioativas (BORGES et al., 2017).

Dentre as famílias com interesse etnomedicinal, a família Lamiaceae é uma das mais importantes possuindo uma grande variedade de plantas com aplicações nas indústrias de cosméticos, nutracêuticos e de medicamentos (ETSASSALA et al., 2021; URITU et al., 2018). É conhecida por apresentar uma grande diversidade de metabólitos secundários, dentre eles, estão os diterpenos que são produzidos abundantemente. Estes por sua vez, conferem várias atividades farmacológicas, como exemplo temos, antimicrobiana (RADULOVIĆ; DENIĆ; STOJANOVIĆ-RADIĆ, 2010), antifúngica (MSONTHQ, 2000), inseticida (RODRÍGUEZ et al., 1993), antioxidante (NDJOUBI et al., 2021), antitumoral (FUJITA et al., 1988), dentre outras.

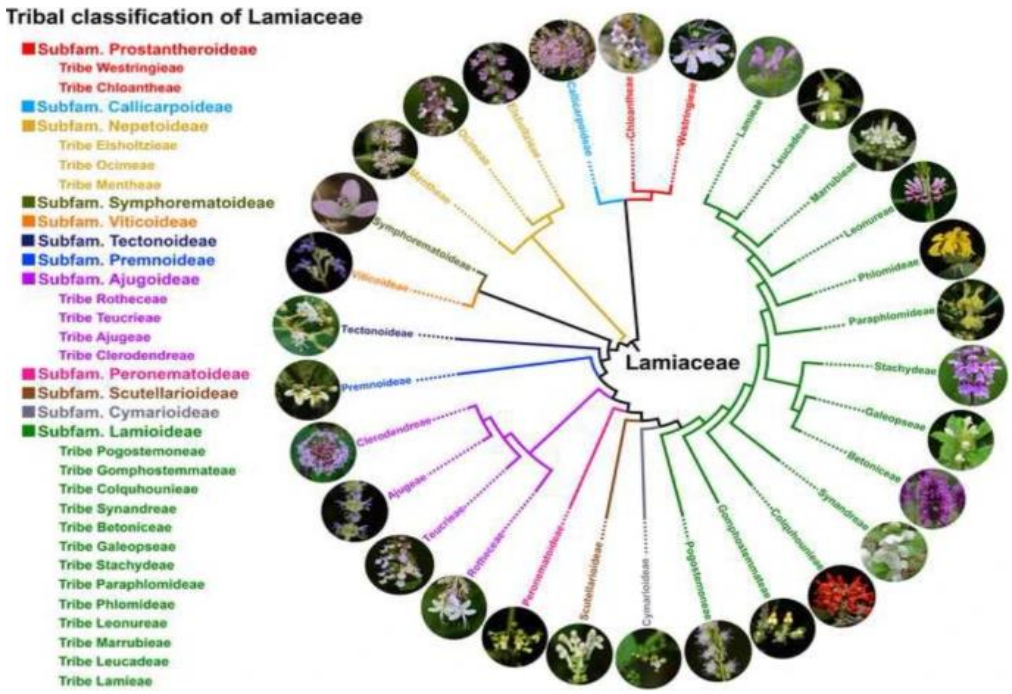
Portanto, reconhecendo a importância da família Lamiaceae e o seu extenso uso popular buscou-se realizar a triagem virtual dos diterpenos da família Lamiaceae, selecionar os mais promissores para realizar atividade biológica. Assim, contribuindo para o planejamento de compostos derivados de produtos naturais candidatos a novos fármacos, mais eficientes e seguros. E através do estudo quimiotaxonômico auxiliar na prospecção de diterpenos com características estruturais específicas, direcionando a um potencial atividade biológica por ferramentas computacionais.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Considerações sobre a família Lamiaceae

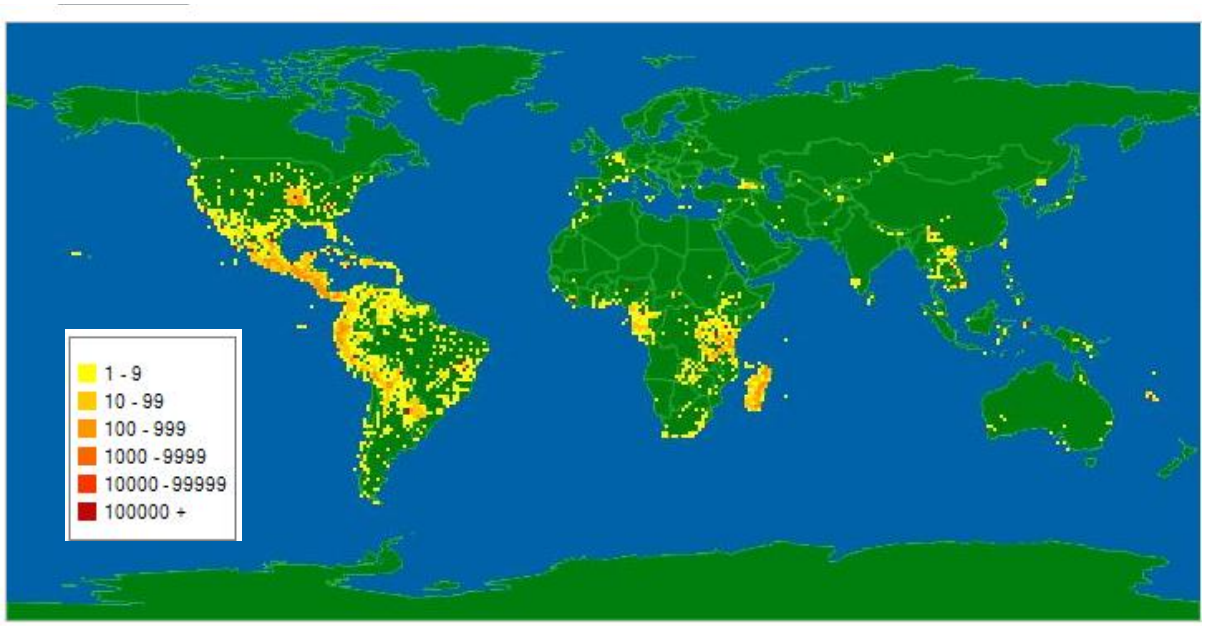
É conhecida como família da hortelã, apresentando uma grande diversidade de óleos essenciais aromáticos, utilizados como condimentos importantes na culinária, sendo apreciadas pelo aroma ou pelo sabor que comunicam através dos alimentos (FALCÃO; FERNANDES; MENEZES, 2003). As Lamiaceaes são frequentemente utilizadas na medicina tradicional como também possuem uma ampla importância econômica como fontes de madeira (por exemplo, *Tectona grandis* L. f.), paisagens ornamentais (por exemplo, sábio escarlate [*Salvia splendens* Sellow ex Wied-Neuw.]), cosméticos (por exemplo, lavanda [*Lavandula angustifolia* Mill.]), ervas culinárias (por exemplo, manjericão *Ocimum basilicum* L.], orégano [*Origanum vulgare* L.]) (ZHAO et al., 2021). Atualmente, Lamiaceae é o maior clado de nível familiar dentro de Lamiales, uma ordem que compreende 26 famílias e mais de 20.000 espécies (LI et al., 2016). É dividida em 12 subfamílias (Ajugoideae, Lamioideae, Nepetoideae, Prostantheroideae, Scutellarioideae, Symphorematoideae, Viticoideae, Cymarioideae, Peronematoideae, Premnoideae, Callicarpoideae e Tectonoideae) (Figura 1).

Figura 1 – Classificação de tribo e subfamília atualizada de Lamiaceae com base na filogenômica do plastomo (ZHAO et al., 2021).



A família Lamiaceae compreende mais de 230 gêneros e mais de 7.000 espécies sendo a sexta maior família das angiospermas (ZHAO et al., 2021). Suas espécies são representadas em geral por ervas, subarbustos ou arbustos e possuem distribuição cosmopolita (LI et al., 2016) principalmente na região do Mediterrâneo e Ásia Central (LEMES; FERRI; LOPES, 2011). Estima-se que no Brasil existem aproximadamente 36 gêneros e 490 espécies (SOARES; PASTORE; JARDIM, 2019). No mapa observado na Figura 2 está a sua distribuição.

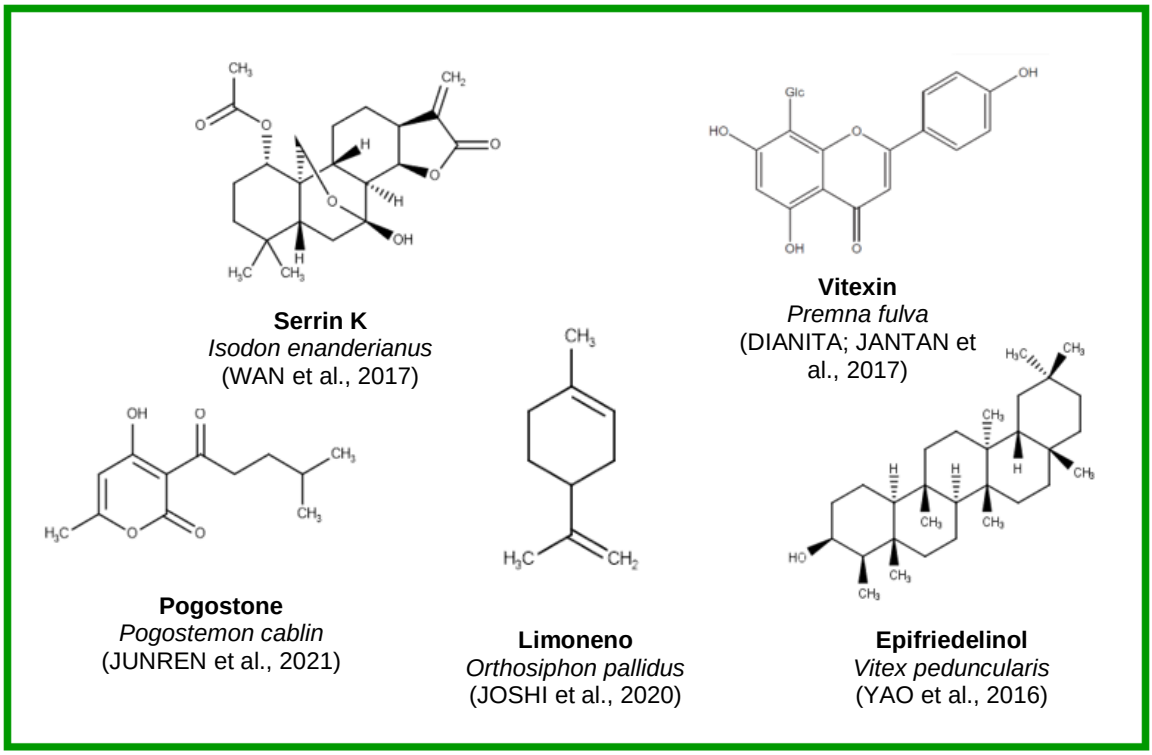
Figura 2 – Mapa de distribuição da família Lamiaceae no mundo.



Fonte: www.tropicos.org/Name/42000291?tab=maps. Acessado em 27 de 10 de 2021)

Em relação ao metabolismo secundário, as espécies desta família concentram uma ampla variedade de estruturas nas quais encontramos diversos marcadores quimiotaxonômicos em todos os níveis: subfamília, gênero e espécie (FREZZA et al., 2019). Dentre esses principais metabólitos secundários característicos dessa família tem-se os monoterpenos (KARAKAYA et al., 2020), sesquiterpenos (JASSBI et al., 2016), diterpenos (HANSON, 2017), triterpenos (YAO et al., 2016), pironas (MARTÍNEZ-FRUCTUOSO et al., 2019), iridoides (DIANITA; JANTAN, 2017), compostos fenólicos (BENEDEC et al., 2015), lignanas (PICKING et al., 2013), flavonoides (SHANG et al., 2010), entre outros (Figura 3).

Figura 3 – Substâncias isoladas de espécies da família Lamiaceae (DIANITA; JANTAN, 2017; JOSHI, 2020; JUNREN et al., 2021; WAN et al., 2017; YAO et al., 2016).

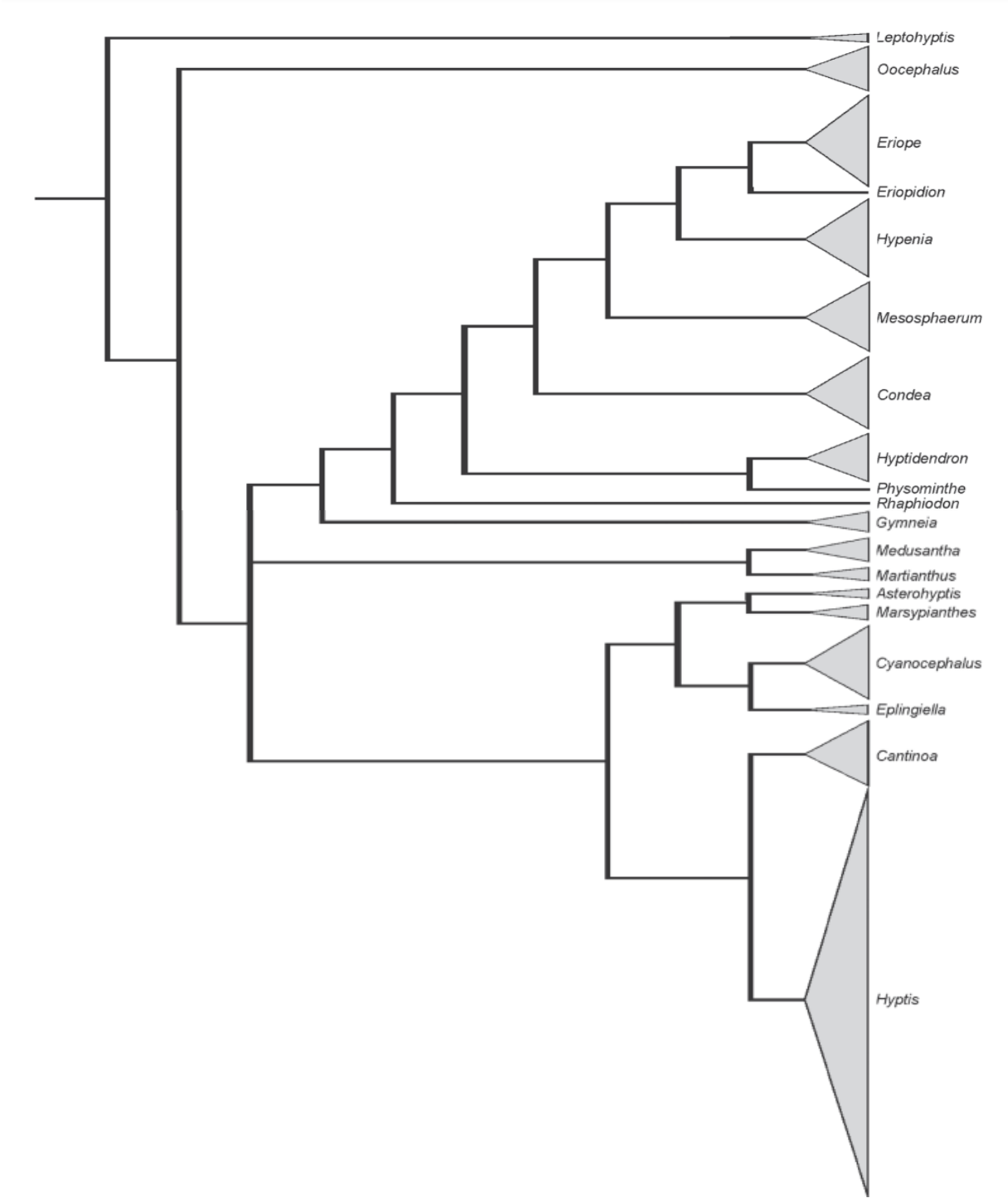


2.2 Subtribo Hyptidinae (Nepetoideae)

A subtribo Hyptidinae possui 19 gêneros e cerca de 400 espécies sendo ervas e arbustos que se encontram distribuídas desde o sul dos Estados Unidos até a Argentina, com ocorrência no Caribe (Figura 4). É composta por plantas que produzem óleos essenciais com odor característicos tendo seu uso popular como repelentes de pragas, no tratamento respiratório e de distúrbios gastrintestinais, entre outros (BRIDI; DE CARVALHO MEIRELLES; VON POSER, 2020).

No Brasil está concentrado a maior parte da diversidade dessa subtribo, sendo *Hyptis* Jacq. seu principal gênero. Diante de estudos morfológicos e moleculares recentes da subtribo Hyptidinae, ocorreu um desmembramento da subtribo em 12 novos gêneros, resultando em 19 gêneros (Figura 4) (BRIDI; DE CARVALHO MEIRELLES; VON POSER, 2020; HARLEY; PASTORE, 2012). Estão incluídos nesse desmembramento algumas espécies dos gêneros, *Leptohyptis* e *Mesosphaerum*, são eles, os gêneros estudados neste trabalho.

Figura 4 – Relações filogenéticas de gêneros de Hyptidinae (HARLEY; PASTORE, 2012).



2.3 *Leptohyptis macrostachys* (Benth.) Harley & J.F.B.Pastore.

Leptohyptis macrostachys (Benth.) Harley & J.F.B.Pastore.≡ *Hyptis macrostachys* Bentham (HARLEY; PASTORE, 2012) é uma espécie arbustiva

pertencente à família Lamiaceae, encontrada no semiárido nordestino e popularmente conhecida como “alfavaca-brava” e “hortelã-do-mato” (Figura 5). Suas folhas são utilizadas na medicina popular no tratamento da asma, tosse e bronquite na forma de infuso ou xarope (AGRA et al., 2008).

Figura 5 – *Leptohyptis macrostachys* (Benth.) Harley & J.F.B.Pastore.

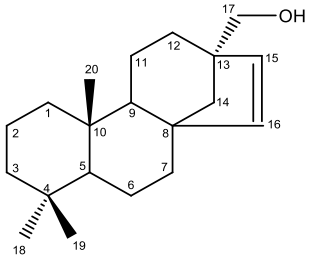
Fonte: Josean F. Tavares, (2009)

O extrato etanólico bruto oriundo das partes aéreas de *L. macrostachys* (LM-EtOH$_{PA}$) apresentou atividade em vários músculos lisos isolados (aorta de rato, útero de rato e traqueia e íleo de cobaia) sendo o maior efeito espasmolítico no íleo de cobaia (COSTA et al., 2014).

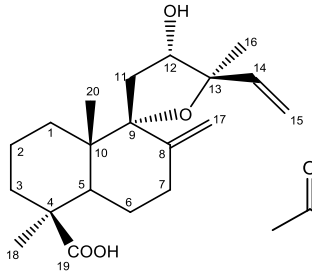
Levantamentos anteriores realizados pelo grupo de pesquisa relataram o estudo fitoquímico em que o extrato etanólico bruto (EEB) e a fase diclorometânica obtidos das folhas e dos caules de *L. macrostachys* resultou no isolamento de dois diterpenos, o eritroxilol B e o hiptenol; dois triterpenos, o acetato de ácido oleanólico e o ácido betulínico; e duas pironas o hiptenolídeo (COSTA, 2013) e 5’-Deacetoxi-hiptenolídeo (SILVA, 2014) dois produtos naturais inéditos (Figura 6). Posteriormente, os estudos farmacológicos reportaram que o (hiptenolídeo), apresentou efeito espasmolítico em íleo de cobaia (COSTA et al., 2014).

Da fase acetato de etila das folhas e dos caules de *L. macrostachys* foram obtidos os triterpenos, ácido tormêntico (COSTA, 2013), ácido oleanólico, acetato de ácido oleanólico e o diterpeno caurano, ácido *ent*-3β-acetoxi-caur-15-eno-17-oico (Figura 6) (SILVA, 2016).

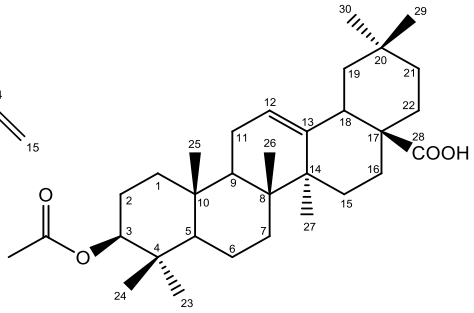
Figura 6 – Substâncias isoladas de *Leptohyptis macrostachys* (Benth.) Harley & J.F.B.Pastore.



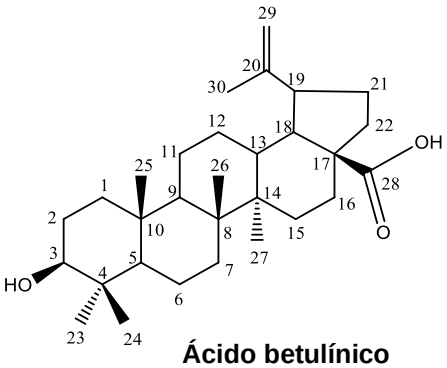
Eritroxilol B
(COSTA, 2013)



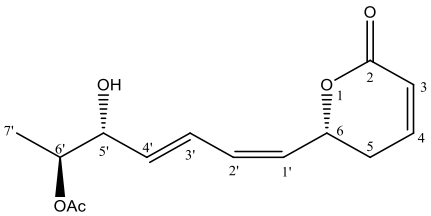
Hiptenol
(COSTA, 2013)



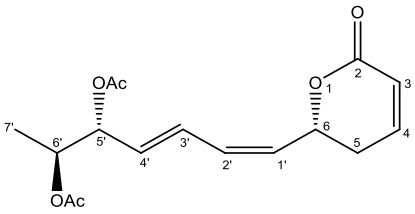
Acetato do ácido oleanólico
(COSTA, 2013)



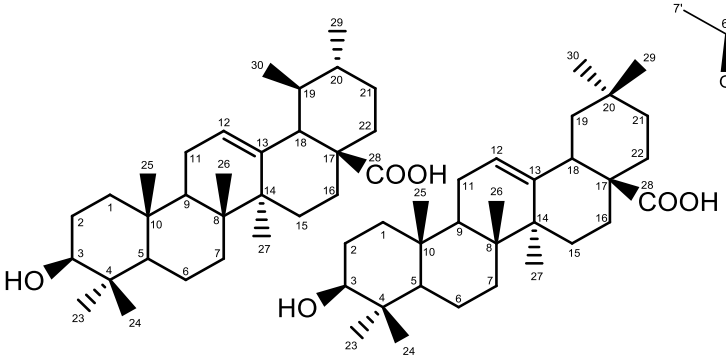
Ácido betulínico
(COSTA, 2013)



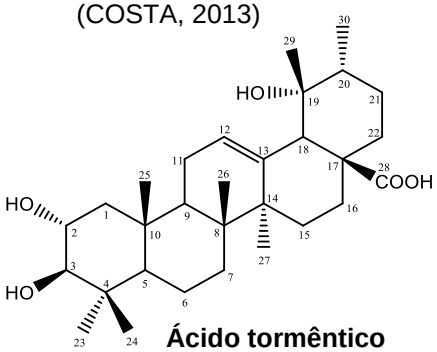
5'-Deacetoxi-hiptenolídeo
(SILVA, 2014)



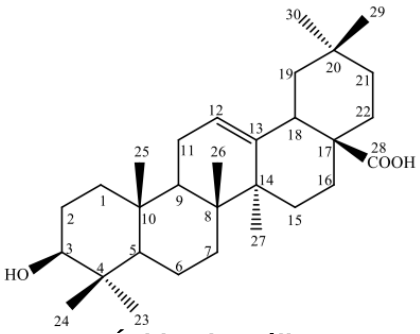
Hiptenolídeo
(COSTA, 2013)



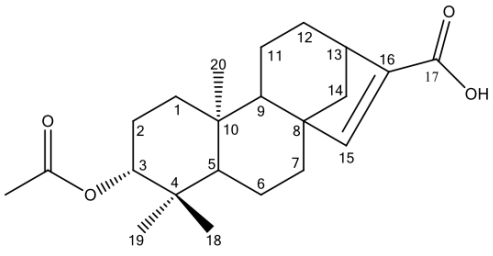
Ácido ursólico + ácido oleanólico
(COSTA, 2013)



Ácido tormêntico
(COSTA, 2013)



Ácido oleanólico
(Silva, 2016)



Ácido *ent*-3β-acetoxi-caur-15-eno-17-óico
(Silva,2016)

2.4 *Mesosphaerum sidifolium* (L'Hérit.) Harley & J.F.B.Pastore.

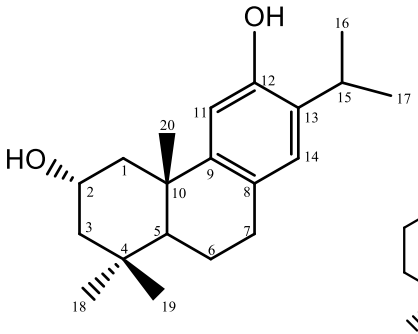
A espécie *Mesosphaerum sidifolium* (L'Hérit.) Harley & J.F.B.Pastore = *Hyptis umbrosa* Salzm. ex Bentham (HARLEY; PASTORE, 2012) também pertence à família Lamiaceae, é conhecida popularmente como “bamburral” (AGRA et al., 2008) e “Aleluia de serrote” (SILVA et al., 2004) sendo uma planta subarbustiva com hábito perene/anual (Figura 7). Na medicina popular as partes mais utilizadas dessa espécie são as folhas e as flores, geralmente para tratamento das doenças de estômago, nasais, articulares, como também funções expectorantes (AGRA et al., 2008).

Figura 7 – Espécie *Mesosphaerum sidifolium* (L'Hérit.) Harley & J.F.B.Pastore.

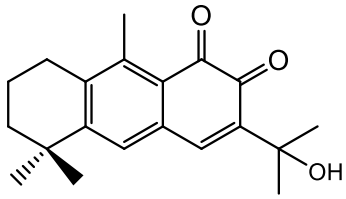
Fonte: Josean F. Tavares, (2009).

De acordo coma literatura, existem poucos estudos com essa espécie, tanto do ponto de vista químico quanto farmacológico. Sendo assim os constituintes químicos encontrados nesta espécie foram os triterpenos: ácido ursólico e 2α-hidroxi-ursólico (MATIDA; ZELNIK; PANIZZA, 1986); os diterpenos: umbrosona (DELLE MONACHE et al., 1990), Pomiferin D, Salviol (COSTA, 2013), pomiferin E, 2α-hidroxisugiol e derivado esterificado do Pomiferin D; os compostos fenólicos: ácido rosmarínico e ácido cafeico (Figura 8) (SILVA, 2016). O extrato etanólico de *Mesosphaerum sidifolium* apresenta uma significante atividade antimicrobiana (DELLE MONACHE et al., 1990b). Também observou que o óleo essencial de suas folhas apresenta potente atividade antitumoral e baixa toxicidade frente aos eritrócitos murinos (CH_{50} = 494,9 µg/ml) e baixa toxicidade aguda em camundongos por via intraperitoneal (DL_{50} em torno de 500 mg/kg) (ROLIM et al., 2017).

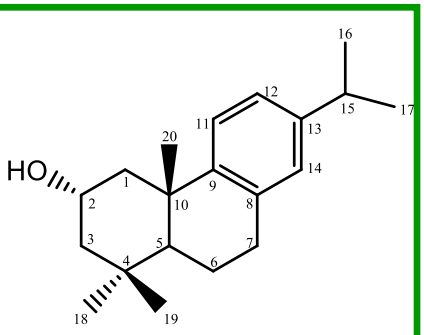
Figura 8 – Substâncias isoladas de *Mesosphaerum sidifolium* (L'Hérit.) Harley & J.F.B.Pastore.



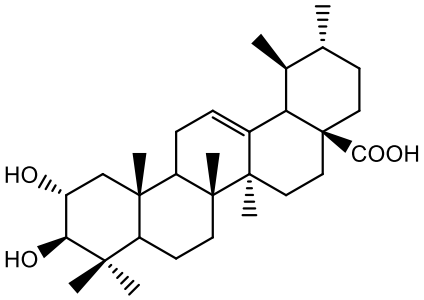
Pormiferin D
(COSTA, 2013)



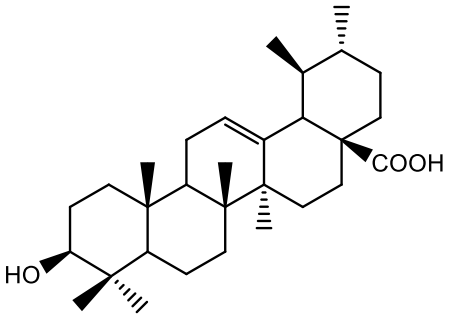
Umbrosona
(DELLE MONACHE et al., 1990)



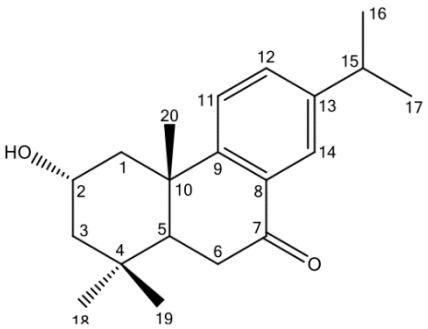
Salviol
(COSTA, 2013)



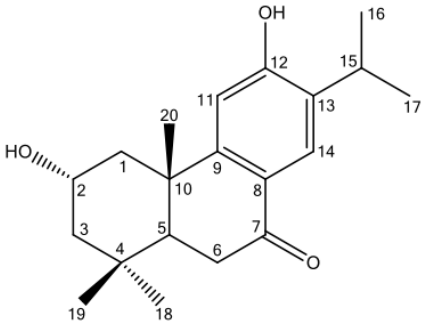
Ácido 2α-hidroxi-ursólico
(MATIDA; ZELNIK; PANIZZA, 1986)



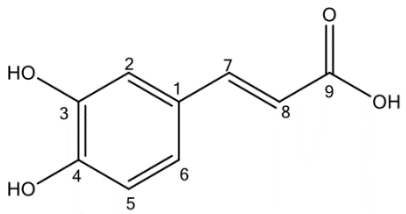
Ácido ursólico
(MATIDA; ZELNIK; PANIZZA, 1986)



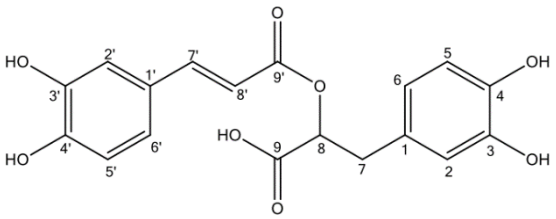
Pomiferin E
(Silva, 2016)



2α-hidroxisugiol
(Silva, 2016)



Ácido cafeico
(Silva, 2016)



Ácido Rosmarínico
(Silva, 2016)

2.5 Terpenoides

As plantas produzem uma ampla quantidade de compostos com estruturas químicas complexas que são classificadas como metabólitos secundários (SIMÕES et al., 2016). Esses por sua vez possuem grande relevância nos processos de polinização, visto que, são responsáveis pela produção de atraentes de polinizadores em flores como também contribuem para a resistência dos organismos pela defesa contra pestes e outras doenças. Consequentemente, determinam também importantes propriedades na qualidade de alimentos utilizados pelo homem, tais como sabor, coloração e aroma, e assim consegue manter a sobrevivência nas condições ambientais favoráveis (BRAZ FILHO, 2010).

Dentre os metabolitos secundários que apresentam tais propriedades, estão os terpenoides que se apresentam bastante numerosos e estruturalmente diversos. Quimicamente, os terpenos são derivados da unidade de isopreno, o 2-metilbuta-1,3-dieno (C_5H_8). Sua classificação é baseada no número e na organização estrutural dos carbonos formados pelo arranjo linear de unidades de isopreno seguidas por condensação de um número de rearranjos do esqueleto de carbono (SIMÕES et al., 2016).

A biossíntese desses metabólitos pode ser dividida em duas rotas principais a via ácido mevalônico pela condensação de três moléculas de acetil-CoA e descarboxilação subsequente e pela via alternativa de blocos de construção C-5, que, a partir da glicose, leva à formação de dois intermediários-chave – 1-desóxi-D-xilulose (DOX) e 2-metileritritol fosfato (MEP) –, que vão originar, em última instância, IPP e DMAPP (Figura 9). A partir dessa via são obtidos tanto blocos de construção de hemiterpenos (C-5) quanto de monoterpenos (C-10), diterpenos (C-20) e tetraterpenos (C-40) (Figura 10). As estruturas contêm esqueletos de carbonos representado por $(C_5)n$, e são classificados como hemiterpenos (C_5), monoterpenos (C_{10}), sesquiterpenos (C_{15}), diterpenos (C_{20}), Sesterterpenos (C_{25}), triterpenos (C_{30}) e triterpenos (C_{40}) (DEWICK, 2002).

Figura 9 – As rotas biossintéticas para a síntese de metabólitos especiais ocorrem pelas seguintes vias: 1 via dos aminoácidos; 2 via do chiquimato; 3 via dos policetídeos; 4 via do ácido mevalônico; e 5 via do metileritritol fosfato (SIMÕES et al., 2016).

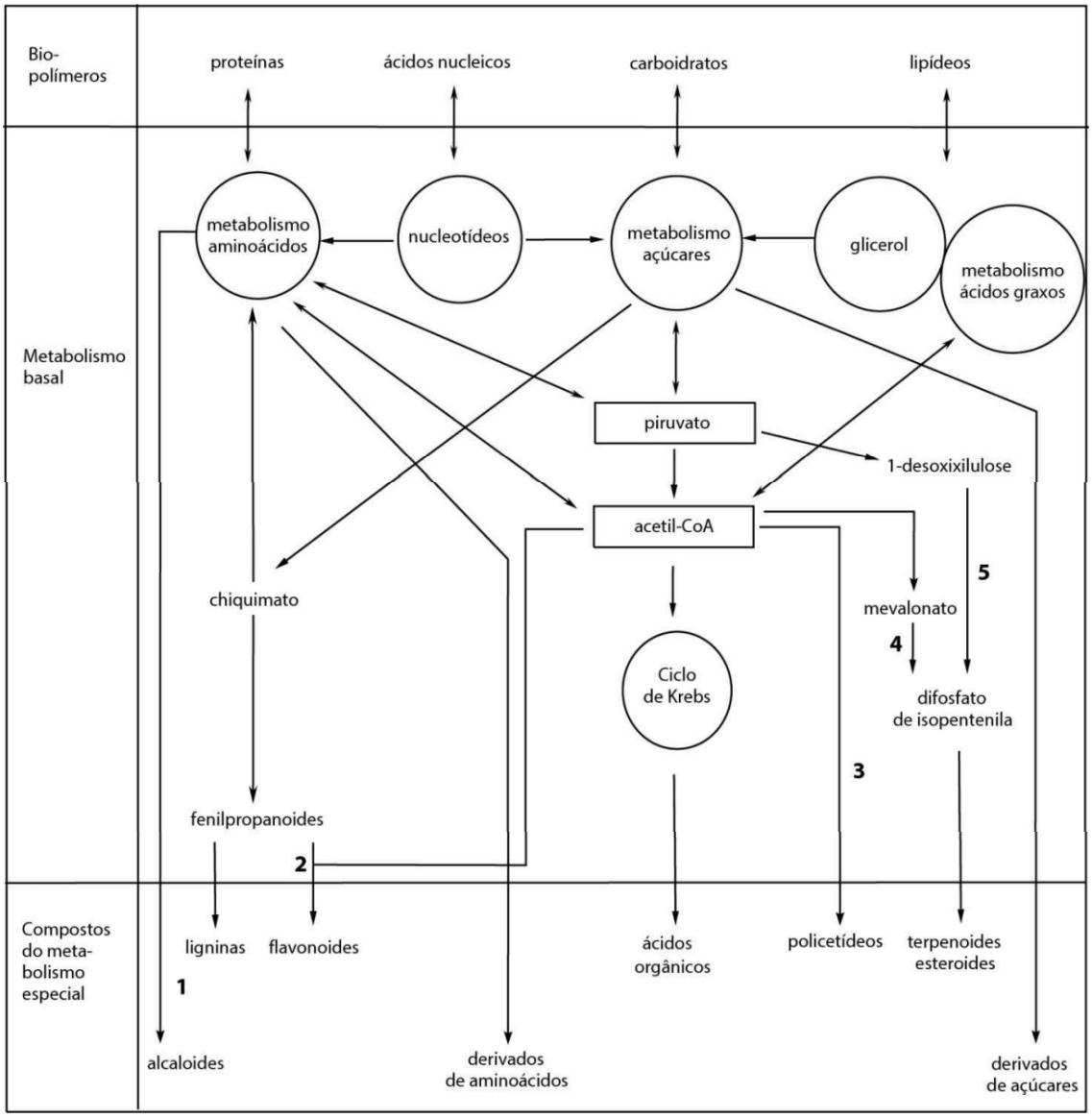
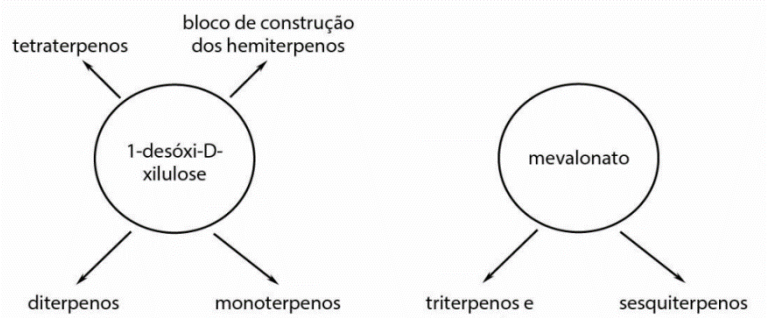


Figura 10 – Origem biossintética dos terpenoides, adaptado de (SIMÕES et al., 2016).



2.6 Triagem virtual e Banco de dados

A descoberta de novas substâncias candidatas a fármacos é um processo de desenvolvimento altamente oneroso. Sendo assim através da descoberta é possível sugerir novas estruturas químicas de fármacos utilizando a etnofarmacologia, química de produtos naturais, triagem de bibliotecas químicas entre outros (AMARAL et al., 2017; BORGES et al., 2017). As técnicas de desenho assistido por computador (CADD) são amplamente utilizadas para reduzir custos e o tempo envolvido no design de drogas (SIMÕES et al., 2018).

No processo para obtenção de um novo fármaco é possível aplicar-se inúmeras ferramentas computacionais de quimioinformática, que desempenham um papel indispensável no design moderno de medicamentos. Neste contexto, como etapa inicial para esse desenvolvimento prioriza-se a identificação de possíveis ligantes para um alvo biológico enquanto que nas etapas finais trabalha-se focado nos ligantes que possuem propriedades físico-químicas, farmacocinéticas e farmacodinâmicas ideais, para que finalmente um fármaco possa adentrar no mercado farmacêutico (RODRIGUES et al., 2012). Para tanto, atualmente, a triagem virtual caracteriza-se por ser uma ferramenta rápida e de baixo custo para a triagem e seleção de *hits*, reduzindo assim a seleção de moléculas classificadas para serem avaliadas experimentalmente (ALVES et al., 2017).

Atualmente a triagem virtual é considerada uma das principais técnicas usadas na descoberta de novos protótipos, pois utiliza métodos computacionais que auxiliam na escolha de moléculas orgânicas promissoras como ligantes de alvos terapêuticos de interesse (PÉREZ-SIANES; PÉREZ-SÁNCHEZ; DÍAZ, 2018a; SHAN et al., 2020). Essas moléculas candidatas clínicas selecionadas por detecção de drogas devem ter um perfil respondendo a diferentes critérios, que se baseiam não apenas na potência do efeito, mas também na seletividade, segurança e também nas chamadas propriedades ADMET (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade) (MENDOLIA et al., 2020).

A triagem virtual utiliza métodos computacionais e são geralmente divididos em métodos baseados no ligante (LBVS, do inglês, *Ligand-based Virtual Screening*) e na estrutura (SBVS, do inglês, *Structure-based Virtual Screening*). Os métodos

LBVS baseiam-se no princípio de que "ligantes semelhantes se ligam de maneira semelhante", exemplificando esse grupo tem-se as impressões digitais moleculares e os métodos de comparação de formas, que possuem a finalidade de avaliar uma biblioteca de moléculas em relação a uma molécula de referência ou uma coleção de ligantes conhecidos.

QSAR por sua vez, permite que a química dos compostos seja transformada em números reprodutíveis, possibilitando assim o uso de orientação quantitativa específica para projetar novos produtos químicos de interesse com maior potencial de desejabilidade (CHERKASOV et al., 2014). Nos dias de hoje, para se obter os modelos de QSAR utiliza-se uma variedade de moléculas, que são encontradas em bancos de dados, e descritores moleculares que fornecem propriedades destas estruturas químicas. Sendo assim, os modelos de QSAR contribuem para fases de descoberta e avaliação de estruturas químicas, transformando então uma informação química em conhecimento (ALVES et al., 2017).

Entretanto os métodos SBVS utiliza os dados estruturais da proteína considerando a estrutura tridimensional (3D) do alvo terapêutico e envolvem o acoplamento de estruturas ligantes potenciais, que são selecionadas utilizando os cálculos de docagem (*docking*), com características químicas, eletrônicas e estruturais que favorecem interações com o sítio ativo do alvo molecular (H MA et al., 2012). Os métodos mais lentos, baseados em dinâmica molecular, oferecem uma visão mais realista e detalhada do evento de ligação. Posto que a dinâmica simula um sistema ao longo do tempo por meio das equações de movimento de Newton e um campo de força, descrevendo as interações entre os diferentes átomos. Diante disso, esses métodos são capazes de evidenciar informações importantes, como as vias de ligação e desacoplamento que um ligante segue bem como as barreiras de alta energia que ele ultrapassa, ou a descoberta de bolsas alostéricas ou locais crípticos (VARELA-RIAL; MAJEWSKI; DE FABRITIIS, 2021).

O *docking* molecular é uma triagem virtual baseada na estrutura do receptor em que vai realizar a predição de moléculas obtidas em uma biblioteca que serão capazes de interagir com o sítio ligante do alvo-molecular e ordenar estas moléculas de acordo com a sua afinidade pelo sítio receptor, com o objetivo de identificar ligantes promissores com potencial atividade farmacológica. As enzimas são utilizadas como

alvos devido ao seu relevante papel nas vias bioquímicas associadas a doenças no ser humano (ANDRICOPULO; GUIDO; OLIVA, 2008; MESTRES, 2005).

Os compostos naturais disponíveis em bibliotecas são de extrema importância para triagem *in vitro* e *in silico* na identificação de novos fármacos. Essas bases de dados podem ser pesquisadas por estrutura química, dados biológicos, por informações biogeográficas e taxonômicas, composto (novo ou não) em outros indivíduos da mesma espécie, gênero, subfamília e família. Para tanto estão disponíveis bancos de dados que catalogam milhares de moléculas que foram testadas em alguns sistemas biológicos com ou sem efeito. Por conseguinte, essa é uma prática desempenhada por quase todas as grandes empresas farmacêuticas do mundo, logo fornece um ponto inicial para a identificação dos fatores químicos que interferem na atividade biológica e das propriedades físico-químicas responsáveis por uma resposta biológica especifica (DE SOUSA LUIS et al., 2020).

Nos dias de hoje existem diversas bases de dados de moléculas químicas que fornecem informações para diversas áreas. Dentre essas bases disponíveis, tem-se exemplos mostrados na tabela 1 e suas atribuições (ALVES et al., 2017; SCOTTI et al., 2015).

Tabela 1 – Base de dados (ALVES et al., 2018; SCOTTI et al., 2015).

Base de dados	Informações relevantes
ChemSpider (*http://www.chemspider.com/)*	Estruturas das moléculas, propriedades físicas, espectro interativo, referências na literatura e fornecedores de produtos químicos
Chemicalize (*http://chemicalize. com/*)	
SciFinder (*https://scifinder.cas.org/)*	Informações estruturais, reações químicas e publicações científicas e patentes
KnowItAll U (*http://www.knowitallu.com/*)	Comporta 2 milhões de espectros, incluindo infravermelho, ressonância magnética nuclear, massas, etc.
Sistemat X *(https://sistematx.ufpb.br/)*	Estruturas de metabólitos secundários, propriedades físico-químicas e experimentais, tais como dados de RMN, ponto de fusão, tempos de retenção em cromatografia
ChEMBL (*https://www.ebi.ac.uk/ chembl/*)	Informações biológicas para estruturas químicas, como resultados de ensaios *in vitro* e *in vivo*
PubChem (*http://pubchem.ncbi.nlm.nih.gov/*)	
DrugBank (*https://www.drugbank.ca/*)	Fármacos aprovados - estrutura química, propriedades físico-químicas calculadas e experimentais, uso terapêutico e informações mais detalhadas, como propriedades farmacocinéticas, toxicológicas,

	farmacodinâmicas e sobre seu alvo molecular, quando disponíveis
PDB (*Protein Data Bank, https://www.rcsb.org/*)	Proteínas, ácidos nucleicos e outras bio-macromoléculas complexas
BMRDB (*Biological Magnetic Resonance Data Bank, http://www.bmrb.wisc. edu/*)	

2.7 Descritores moleculares

Os descritores moleculares desempenham um grande papel na química, nas ciências farmacêuticas, na política de proteção ambiental, pesquisa em saúde e controle de qualidade, sendo usados para prever as propriedades biológicas e físico-químicas das moléculas (QSAR/QSPR) e para rastreio virtual de bibliotecas de moléculas. Para que uma estrutura química seja compreendida e processada por um computador, ela precisa ser descrita em uma sequência numérica única. Um descritor molecular é o resultado final de um procedimento matemático e lógico que transforma informação química codificada em uma representação simbólica de uma molécula em um número útil, viabilizando sua análise e correlação com características desejáveis (TODESCHINI; CONSONNI, 2009). Os descritores químicos apresentam-se de formas distintas mostrando diferentes níveis de representação estrutural. Esses descritores por sua vez, podem ser classificados quanto à sua “dimensionalidade” (Tabela 2).

Tabela 2 – Classificação de descritores (ALVES et al., 2017).

Representação molecular	Descritores
Unidimensionais (1D)	Baseados em propriedades físico-químicas e da fórmula molecular (ex., peso molecular, refratividade molar, logP, entre outros)
Bidimensionais (2D)	Descrevem propriedades que podem ser calculadas de uma representação 2D (ex., número de átomos, número de ligações, índices de conectividade, entre outros)
Tridimensionais (3D)	Necessitam da conformação 3D das moléculas (ex., volume de Van der Waals, área de superfície acessível ao solvente, entre outros)

Outra classificação está relacionada à natureza desses descritores, podendo ser: constitucionais, que são oriundos da composição atômica do composto (ex., peso molecular, números de átomos e ligações); topológicos (ex., índice de conteúdo de

informações de ligações); geométricos, que são resultantes de coordenadas 3D (ex., volume molecular, área de superfície polar, entre outros); eletrostáticos, que são derivados das cargas parciais (ex., índices de polaridade, carga parciais, entre outros); e mecânico-quânticos, que são gerados das funções de onda dos elétrons (ex., energia dos orbitais moleculares) (ALVES et al., 2017).

O cálculo de descritores pode ser realizado em diversos programas computacionais, como o Dragon 7.0 que calcula 5.270 descritores moleculares, cobrindo a maioria das várias abordagens teóricas. A lista de descritores inclui os tipos de átomos mais simples, contagens de grupos funcionais e fragmentos, descritores topológicos e geométricos, descritores tridimensionais, várias estimativas de propriedades (como logP) e *drug-like* e *lead-like alerts* (como *Lipinski's alert*) (TODESCHINI; CONSONNI, 2009). Sendo organizados em diferentes blocos lógicos que são ainda divididos em sub-blocos para facilitar o gerenciamento, seleção e análise de descritores (VERSION 7.0.10, [s.d.]). A lista com os 30 blocos encontra-se na Tabela 3.

Tabela 3 – Lista dos 30 blocos de descritores moleculares do Dragon 7.0.

Bloco	Nome do Bloco	Nº de Descritores
1	Constitucional	47
2	Descritores de anel	32
3	Índices topológicos	75
4	*Walk and path counts*	46
5	Índices de conectividade	37
6	Índices de informação	50
7	Descritores baseados em matriz 2D	607
8	Autocorrelações 2D	213
9	Autovalores de carga	96
10	Descritores tipo P-VSA	55
11	Índices de ETA	23
12	*Edge adjacency indices*	324
13	Descritores geométricos	38
14	Descritores 3D baseados em matriz	99
15	Autocorrelações 3D	80
16	Descritores RDF	210
17	Descritores 3D-MoRSE	224
18	Descritores WHIM	114
19	Descritores GETAWAY	273
20	Perfis moleculares randômicos	41
21	Contagem de grupos funcionais	154
22	Fragmentos de átomo centrado	115
23	*Atom-type E-state indices*	172
24	CATS 2D	150
25	*2D Atom Pairs*	1596
26	*3D Atom Pairs*	36
27	Descritores de carga	15

28	Propriedades moleculares	20
29	*Drug-like indices*	28
30	CATS 3D	300

2.8 Aprendizado de máquina

O aprendizado de máquina, em inglês *machine learning*, é uma área da Inteligência Artificial que fornece um treinamento automático de funções de pontuação orientadas por dados, ou seja, uma elaboração de sistemas que podem aprender com dados. O aprendizado automático explora o estudo e construção de algoritmos que podem aprender a partir de seus erros e predizer dados. Estes algoritmos operam construindo um modelo a partir de entradas (*inputs)* amostrais a fim de fazer previsões ou decisões guiadas pelos dados (SCHNEIDER; DOWNS, 2003; VARELA-RIAL; MAJEWSKI; DE FABRITIIS, 2021).

Os métodos de aprendizado são divididos em supervisionados e não supervisionados. No supervisionado, o modelo é treinado a partir de uma base de dados com as classes conhecidas previamente, conjunto de treinamento, que é utilizado na criação do modelo, durante a fase de obtenção das regras de classificação. Existem vários algoritmos de aprendizado supervisionados estão disponíveis, como por exemplo o *Randon Forest* (RF), *Support Vector Machine* (SVM), redes neurais (NN) (BADILLO et al., 2020; MITCHELL B.O., 2014). Os modelos resultantes da utilização de métodos supervisionados, uma vez validados, constituem um ponto inicial para a avaliação e/ou seleção de compostos químicos que carecem de dados experimentais (ALVES et al., 2017).

Os métodos não supervisionados são usados para identificar padrões nos conjuntos de dados com base apenas nos descritores, visto que nenhuma informação é dada ao algoritmo de aprendizado, apenas os dados de entrada são apresentados e o algoritmo tem que descobrir sozinho as relações, padrões, regularidades ou categorias nos dados que lhe vão sendo apresentados e codificá-las nas saídas. Quando aplicada aos conjuntos de dados químicos, identifica subgrupos homogêneos entre um conjunto de dados heterogêneo. Alguns dos vários usos dessa abordagem incluem a verificação da diversidade estrutural do conjunto de dados, a avaliação da consistência de dados experimentais e a exploração de possíveis interferências que

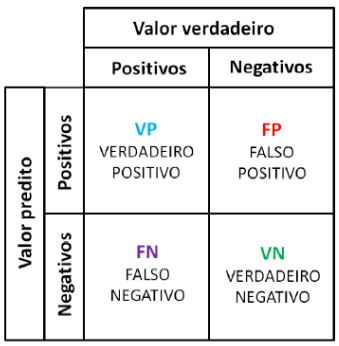
influenciam na atividade. Algoritmos comumente utilizados nessa abordagem incluem análise de componentes principais (PCA, do inglês *principal component analysis*), análise de agrupamentos hierárquicos (HCA, do inglês *hierarchical cluster analysis*) e mapas auto-organizáveis (SOM, do inglês *self-organizing maps*) (ALVES et al., 2017; BADILLO et al., 2020).

2.9 Predição de Dados

2.9.1 Matriz de confusão

É responsável por classificar todos os dados do modelo em categorias, determinando se o valor previsto correspondeu ao valor real. Mostrando a quantidade de dados dos conjuntos de treinamento e teste que foi classificada correta e incorretamente pelo modelo. As linhas na matriz representam os valores previstos para o modelo, sendo que as colunas representam os valores reais. Considerando um modelo classificador de duas classes: Positivos (**P**) e Negativos (**N**), por exemplo, compostos ativos (**P**) e compostos inativos (**N**), as categorias usadas na análise são **Verdadeiros Positivos, Falsos Positivos**, **Verdadeiros Negativos** e **Falsos Negativos**. É considerado **Verdadeiro Positivo**, quando a instância é positiva e é classificada como positiva; se é classificada como negativo, conta-se como **Falso Negativo**. Se a instância é negativa e é classificada como negativa, conta-se como **Verdadeiro Negativo**; se é classificada como positiva, é considerada como **Falso Positivo** (BADILLO et al., 2020; HANLEY; MCNEAL, 1982) (Figura 11).

Figura 11. Matriz de confusão.



A partir dos resultados da matriz de confusão podem ser calculadas as seguintes métricas para determinar as características de cada um dos modelos de predição (BADILLO et al., 2020):

a) **ACURÁCIA:** proporção de predições corretas, sem levar em consideração o que é positivo e o que é negativo.

$$A = \frac{VP + VN}{P + N}$$

b) **SENSIBILIDADE**: proporção de verdadeiros positivos - a capacidade do sistema em predizer corretamente a condição para casos que realmente a têm.

$$S = \frac{VP}{(VP + FN)}$$

c) **ESPECIFICIDADE**: proporção de verdadeiros negativos - a capacidade do modelo em predizer corretamente que um dado é negativo.

$$E = \frac{VN}{(VN + FP)}$$

d) **VERDADEIRO PREDITIVO POSITIVO:** proporção de verdadeiros positivos em relação a todas as predições positivas.

$$VPP = \frac{VP}{(VP + FP)}$$

e) **VERDADEIRO PREDITIVO NEGATIVO**: proporção de verdadeiros negativos em relação a todas as predições negativas.

$$VPN = \frac{VN}{(VN + FN)}$$

2.9.2 Coeficiente de correlação de Matthews

O coeficiente de correlação de Matthews (MCC, do inglês *Matthews correlation coefficient*) é uma medida de qualidade de duas classificações binárias que pode ser utilizada para avaliar de modo global o modelo a partir dos resultados obtidos da matriz de confusão. Apresenta um valor entre (-1) e (+1), em que um coeficiente de (+1) representa uma predição perfeita, (0) representa uma predição aleatória media, e (-1)

indica desacordo total entre previsão e observação (MATTHEWS, 1975). O coeficiente de correlação de Matthews pode ser calculado a partir da seguinte formula:

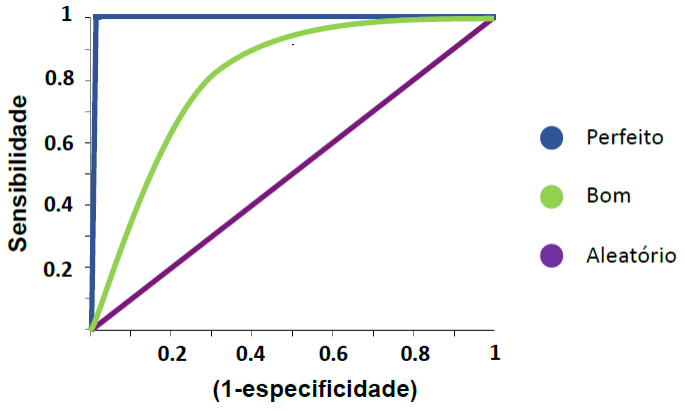
$$MCC = \frac{VP\ x\ VN - FP\ x\ FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

Onde, VP = taxa verdadeiro positivo, VN = taxa verdadeiro negativo, FP = taxa de falso positivo e FN =Taxa falso negativo.

2.9.3 Área Abaixo da Curva ROC

Uma das principais formas de avaliar um classificador é através das curvas ROC. Análise ROC do inglês *Receiver Operating Characteristic*, é a forma mais eficiente de demonstrar a performance de um modelo, através da avaliação, organização e seleção de sistemas de diagnóstico e/ou predição. É um gráfico de sensibilidade (ou taxa de verdadeiros positivos) versus taxa de falsos positivos (1 – especificidade), um método utilizado é através do cálculo da área abaixo da curva ROC (AUC) (HAND, 2001; VUK; CURK, 2006) (Figura 12). A área sob a curva, (AUC) do inglês *Area Under The Curve*, determina estatisticamente a acurácia do teste. Quando uma variável que está a ser investigada não consegue distinguir entre dois grupos, a área sob a curva é igual a 0,5 e a curva ROC coincide com a diagonal. Quando há uma separação perfeita dos valores entre dois grupos, a área sob a curva é igual a 1, então a curva ROC atingirá o canto superior esquerdo do gráfico (VUK; CURK, 2006).

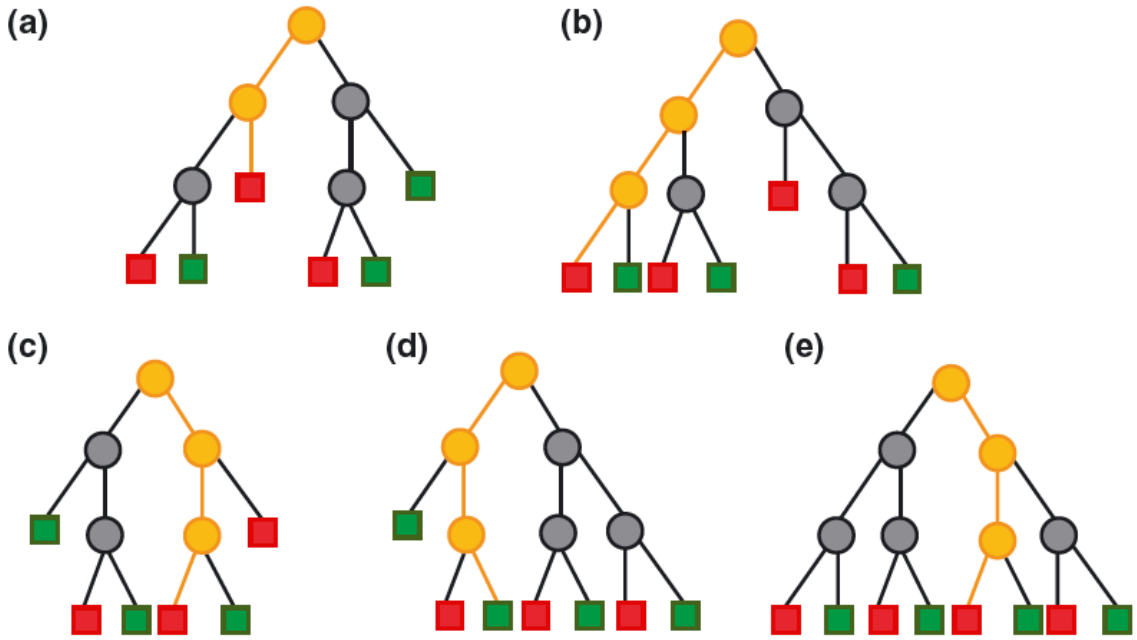
Figura 12 – Curva ROC.



2.10 *Randon Forest*

O *Randon Forest* é uma técnica de classificação baseada em um conjunto, ou floresta, de árvores de decisão. Para determinar a classe de uma instância, o método combina o resultado de várias árvores de decisão, por meio de um mecanismo de votação. Com a quebra das massas de dados e construção de vários subconjuntos, uma árvore de decisão é construída. Sendo assim a construção das árvores ocorre pela seleção de atributos aleatoriamente a partir dos subconjuntos, onde os mesmos são aplicados nos nós de cada uma das árvores criadas. Posteriormente, resulta que em cada árvore dá uma classificação, ou um voto para uma classe. A classificação final é dada pela classe que recebeu o maior número de votos entre todas as árvores da floresta (Figura 13) (DINIZ et al., 2013; MITCHELL B.O., 2014).

Figura 13 – Ilustração da lógica do algoritmo *Randon Forest* - os nós da folha terminal são mostrados como quadrados e coloridos em vermelho ou verde de acordo com a classe. O caminho percorrido por cada árvore por uma instância de consulta é mostrado em laranja. As árvores A, B, C e E preveem que a instância pertence à classe vermelha, árvore D dissidente, de modo que a Randon Forest irá atribuí-la à classe vermelha por um voto majoritário de 4-1 (MITCHELL B.O., 2014).



2.11 Redes Neurais Artificiais e Mapas Auto-Organizáveis

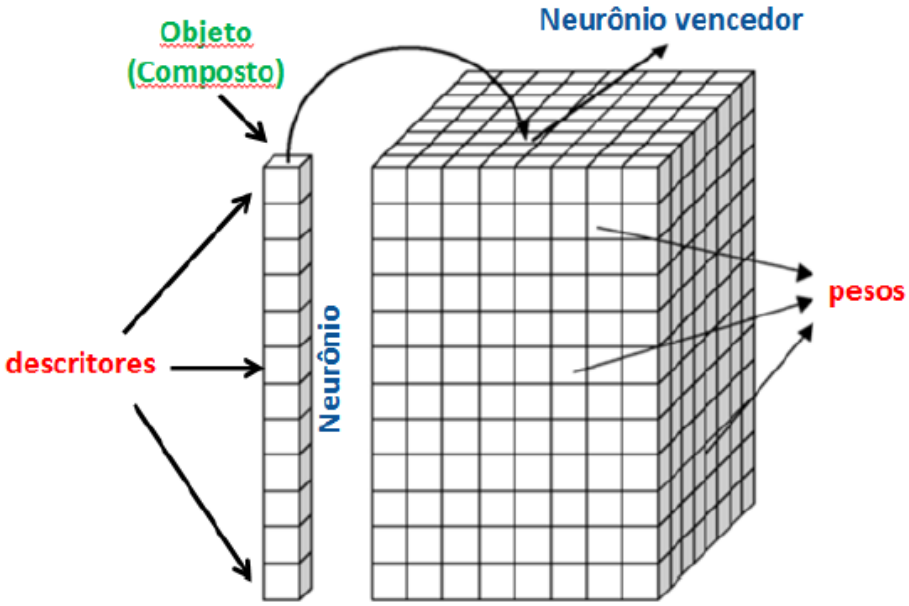
As Redes Neurais Artificiais (RNAs) são técnicas computacionais que demonstram um modelo matemático utilizado para reconhecimento de padrões e aprendizado de máquina inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. A arquitetura da rede é baseada em neurônios conectados em uma camada de entrada, uma camada ou camadas ocultas e uma camada de saída. Cada conexão entre os neurônios comporta um peso que variam durante a fase de treinamento, à medida que a rede aprende como conectar os dados de entrada e saída, antes de ser testada em instâncias invisíveis (MITCHELL B.O., 2014; PÉREZ-SIANES; PÉREZ-SÁNCHEZ; DÍAZ, 2018b).

Mapa auto organizável ("*Self-Organizing Map*" - SOM), também conhecido por Rede Neural Artificial de Kohonen, é um modelo de redes neurais competitivas em maior uso na atualidade, e tem sido utilizado em uma gama de aplicações, incluindo agrupamento e visualização de dados multidimensionais. A rede SOM é usada para mapear um espaço de entrada dimensional, contendo *n* padrões, para uma grade de neurônios uni- ou bidimensional, com os objetivos de quantizar o espaço de entrada e representar da melhor forma possível a topologia original em um espaço de menor dimensão. Uma vez treinado, podemos rotular neurônios caso tenhamos, além da informação dos padrões, as suas respectivas classes. Padrões são apresentados ao mapa e o neurônio vencedor será o mais similar, ou o mais próximo, de acordo com o critério de similaridade escolhido (COSTA; ANDRADE NETTO, 2007).

A arquitetura de um SOM pode ser visualizada através da Figura 14. Cada coluna na grade representa um neurônio, cada caixa dessa coluna representa um peso (um valor numérico) e os objetos a serem mapeadas na rede. Antes do treinamento os pesos tomam valores aleatórios, depois ocorre um ajuste desses valores. Quando um objeto é apresentado à rede, todos os neurônios competem para serem estimulados, mas apenas um neurônio é considerado o neurônio vencedor: aquele com os pesos mais parecidos com os descritores de entrada. Por conseguinte, a rede então corrige os dos neurônios da vizinhança e os pesos do neurônio vencedor também para que eles se tornem ainda mais parecidos com os descritores de entrada. Visto que a rede foi treinada, cada neurônio é classificado de acordo com a classe dos

objetos que foram mapeados. Através do algoritmo de treinamento, ocorre o agrupamento de objetos representados por descritores similares. Para tanto, esse fluxo, então, é repetido iterativamente com todos os objetos (VARNEK, 2017).

Figura 14 – Arquitetura de um mapa auto organizável adaptado de (VARNEK, 2017).



2.12 Tuberculose

A tuberculose (TB) é uma doença infecciosa considerada das mais prevalentes do mundo e anualmente cerca de 10 milhões de pessoas em todo o mundo desenvolvem TB. É causada principalmente pelo *Mycobacterium tuberculosis* sendo a apresentação mais comum a tuberculose pulmonar. Os sintomas são geralmente inespecíficos e pode desenvolver-se insidiosamente no decorrer de várias semanas ou meses. Os portadores podem apresentar perda de peso, fadiga noturna suores e sintomas respiratórios, como tosse, produção de expectoração, dor no peito e dispneia. Não existem vacinas eficazes disponíveis contra a doença o que desperta interesse da indústria farmacêutica por uma terapia ou profilaxia para a TB (FROESSL; ABDEEN, 2021; HAMIDIEH et al., 2021; TAMBURINI et al., 2021).

A *Mycobacterium smegmatis* é encontrada principalmente no solo, água e secreções seborreicas sendo considerada um organismo não patogênico, embora, em

casos raros, possa causar infecções da pele e dos tecidos moles. Seu crescimento é lento e é comumente utilizada como um modelo adequado para investigar a biologia de microbactérias patógenas como a *Mycobacterium tuberculosis* (CZUBAT et al., 2020; SHARMA et al., 2020).

3 OBJETIVOS

3.1 Objetivo Geral

- Contribuir com o conhecimento químico da família Lamiaceae, através da criação de banco de dados e análise quimiotaxonômica, como também predizer por triagem virtual o potencial farmacológico contra *Mycobaterium tuberculosis* dos diterpenos isolados das espécies *Leptohyptis macrostachys* e *Mesosphaerum sidifolium,* e validar experimentalmente.

3.2 Objetivos Específicos

- Criar um banco de dados com os metabolitos secundários (diterpenos) isolados da família Lamiaceae, para obtenção de um perfil químico;

- Realizar estudos quimiotaxonômicos com os diterpenos de Lamiaceae utilizando descritores moleculares;

- Construir modelos de predição de moléculas potencialmente ativas com atividade antimicrobiana para *Mycobaterium tuberculosis*, utilizando descritores moleculares gerados pelo software Dragon 7.0;

- Realizar estudos de modelagem molecular utilizando proteínas alvo da *Mycobaterium tuberculosis* e como ligantes os diterpenos isolados das espécies *Leptohyptis macrostachys* e *Mesosphaerum sidifolium*;

- Realizar triagem *in vitro* frente a *Mycobaterium Tuberculosis*.

REFERÊNCIAS

AGRA, M. D. F. et al. Survey of medicinal plants used in the region Northeast of Brazil. **Revista Brasileira de Farmacognosia**, v. 18, n. 3, p. 472–508, set. 2008.

ALVES, V. et al. Quimioinformática: uma introdução. **Química Nova**, v. 41, p. 202, 2017.

AMARAL, A. T. et al. A evolução da Química Medicinal no Brasil: avanços nos 40 anos da Sociedade Brasileira de Química. **Química Nova**, v. 40, n. 6, p. 694–700, 3 jul. 2017.

ANDRICOPULO, A.; GUIDO, R.; OLIVA, G. Virtual Screening and Its Integration with Modern Drug Design Technologies. **Current Medicinal Chemistry**, v. 15, n. 1, p. 37–46, 2008.

BADILLO, S. et al. An Introduction to Machine Learning. **Clinical Pharmacology and Therapeutics**, v. 107, n. 4, p. 871–885, 1 abr. 2020.

BENEDEC, D. et al. Assessment of rosmarinic acid content in six Lamiaceae species extracts and their antioxidant and antimicrobial potential. **Pak. J. Pharm. Sci**, v. 28, n. 6, p. 2297–2303, 2015.

BORGES, W. DE S. et al. A Química de Produtos Naturais do Brasil no Século XXI. **Química Nova**, v. 40, n. 6, p. 706–710, 3 jul. 2017.

BRAZ FILHO, R. Contribuição da fitoquímica para o desenvolvimento de um país emergente. **Química Nova**, v. 33, p. 229–239, 2010.

BRIDI, H.; DE CARVALHO MEIRELLES, G.; VON POSER, G. L. Subtribe Hyptidinae: a promising source of bioactive metabolites. **Journal of Ethnopharmacology**, p. 113225, 2020.

CALIXTO, J. B. The role of natural products in modern drug discovery. **Anais da Academia Brasileira de Ciencias**, v. 91, p. 1–7, 2019.

CASANOVA, L. M.; COSTA, S. S. Synergistic Interactions in Natural Products: Therapeutic Potential and Challenges. **Revista Virtual de Química**, v. 9, n. 2, p. 575–595, 2017.

CHERKASOV, A. et al. QSAR Modeling: Where Have You Been? Where Are You Going To? **Journal of Medicinal Chemistry**, v. 57, n. 12, p. 4977–5010, 26 jun. 2014.

COSTA, J. A. F.; ANDRADE NETTO, M. L. DE. Segmentação de mapas auto-organizáveis com espaço de saída 3-D. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, v. 18, p. 150–162, 2007.

COSTA, V. C. DE O. et al. Hyptenolide, a new α-pyrone with spasmolytic activity from *Hyptis macrostachys*. **Phytochemistry letters**, v. 8, p. 32–37, 2014.

COSTA, V. C. DE OLIVEIRA. Contribuição ao conhecimento químico de espécies de Hyptis com ocorrência no semiárido paraibano: *Hyptis macrostachys Benth.* e *Hyptis umbrosa* Salzm. ex Benth. **Tese**, p. 1 a 162, João Pessoa: 2013.

CZUBAT, B. et al. Functional Disassociation Between the Protein Domains of MSMEG_4305 of *Mycolicibacterium smegmatis* (*Mycobacterium smegmatis*) in vivo. **Frontiers in Microbiology**, v. 11, 19 ago. 2020.

DE SOUSA LUIS, J. A. et al. Virtual screening of natural products database. **Mini-Reviews in Medicinal Chemistry**, v. 20, 3 ago. 2020.

DELLE MONACHE, F. et al. Umbrosone, an ortho-quinone from *Hyptis umbrosa*. **Phytochemistry**, v. 29, n. 12, p. 3971–3972, 1990a.

DEWICK, P. M. **Medicinal natural products: a biosynthetic approach**. [s.l.] John Wiley & Sons, 2002.

DIANITA, R.; JANTAN, I. Ethnomedicinal uses, phytochemistry and pharmacological aspects of the genus Premna : a review. **Pharmaceutical Biology**, v. 55, n. 1, p. 1715–1739, 9 jan. 2017.

DIAS, D. A.; URBAN, S.; ROESSNER, U. A Historical Overview of Natural Products in Drug Discovery. **Metabolites**, v. 2, n. 2, p. 303–336, 16 abr. 2012.

DINIZ, F. A. et al. RedFace: um sistema de reconhecimento facial baseado em técnicas de análise de componentes principais e autofaces. **Revista Brasileira de Computação Aplicada**, v. 5, n. 1, p. 42–54, 2013.

DRUMOND, M. A.; SCHISTEK, H.; SEIFFARTH, J. A. Caatinga, ecossistema heterogêneo. **Revista do Instituto Humanitas Unisinos**, v. 389, p. 1–60, 2012.

ETSASSALA, N. G. E. R; Hussein, A. A; Nchu, F. Potential Application of Some Lamiaceae Species in the Management of Diabetes. **Plants**, v. 10, 279, 2021.

FALCÃO, D. Q.; FERNANDES, S. B. O.; MENEZES, F. S. Triterpenos de *Hyptis fasciculata* Benth. **Revista Brasileira de Farmacognosia**, v. 13, p. 81–83, 2003.

FREZZA, C. et al. Phytochemistry, Chemotaxonomy, Ethnopharmacology, and Nutraceutics of Lamiaceae. In: **Studies in Natural Products Chemistry**. 1. ed. [s.l.] Elsevier B.V., 2019. v. 62p. 125–178.

FROESSL, L. J.; ABDEEN, Y. Pseudomembranous Tracheobronchitis due to *Mycobacterium tuberculosis*. **Cureus**, v. 13, n. 8, 2021.

FUJITA, T. et al. Cytotoxic and Antitumor Activities of Rabdosia Diterpenoids. **Planta Medica**, v. 54, n. 05, p. 414–417, 1988.

H MA, X. et al. Virtual screening methods as tools for drug lead discovery from large chemical libraries. **Current medicinal chemistry**, v. 19, n. 32, p. 5562–5571, 2012.

HAMIDIEH, F. et al. An overview of genetic information of latent *mycobacterium tuberculosis* and Respiratory Diseases**. Korean National Tuberculosis Association**, v. 84, p. 1-12, 2021.

HAND, D. J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. **Springer**, v. 45, p. 171-186, 2001.

HANLEY, J. A.; MCNEAL, B. J. A simple generalization of the area under the ROC curve to multiple class classification problems. **Radiology**, v. 143, p. 29–36, 1982.

HANSON, J. R. Diterpenoids of terrestrial origin. **Natural Product Reports**, v. 34, n. 10, p. 1233–1243, 2017.

HARLEY, R. M.; PASTORE, J. F. B. A generic revision and new combinations in the Hyptidinae (Lamiaceae), based on molecular and morphological evidence. **Phytotaxa**, v. 58, p. 1–55, 2012.

JASSBI, A. R. et al. Bioactive phytochemicals from shoots and roots of *Salvia* species. **Phytochemistry Reviews**, v. 15, n. 5, p. 829–867, 2016.

JOSHI, R. K. GC-MS analysis of the volatile constituents of *Orthosiphon pallidus* Royle, ex Benth. **Natural Product Research**, v. 34, n. 3, p. 441–444, 1 fev. 2020.

JUNREN, C. et al. Pharmacological activities and mechanisms of action of *Pogostemon cablin* Benth: a review. **Chinese Medicine (United Kingdom)**BioMed Central Ltd, , 1 dez. 2021.

KARAKAYA, S. et al. A caryophyllene oxide and other potential anticholinesterase and anticancer agent in *Salvia verticillata* subsp. *amasiaca* (Freyn & Bornm.) Bornm. (Lamiaceae). **Journal of Essential Oil Research**, v. 32, n. 6, p. 512–525, 1 nov. 2020.

LEMES, G. DE F.; FERRI, P. H.; LOPES, M. N. Constituintes químicos de *Hyptidendron canum* (Pohl ex Benth.) R. Harley (Lamiaceae). **Química Nova**, v. 34, n. 1, p. 39–42, 2011.

LI, B. et al. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. **Scientific Reports**, v. 6, n. 1, p. 34343, 17 out. 2016.

MAMADALIEVA, N. Z. et al. The genus *lagochilus* (Lamiaceae): A review of its diversity, ethnobotany, phytochemistry, and pharmacology**. Plants,** v. 10, n. 1, p. 132, 2021.

MARTÍNEZ-FRUCTUOSO, L. et al. Structure Elucidation, Conformation, and Configuration of Cytotoxic 6-Heptyl-5,6-dihydro-2 H -pyran-2-ones from *Hyptis* Species and Their Molecular Docking to α-Tubulin. **Journal of Natural Products**, v. 82, n. 3, p. 520–531, 22 mar. 2019.

MATIDA, A. K.; ZELNIK, R.; PANIZZA, S. Presença dos ácidos ursólicos e 2α-hidroxi-ursólico *Hyptis umbrosa* Salzmann e *Eriope crassipes* Bentham (Labiatae). **REVISTA BRASILEIRA DE FARMACOGNOSIA**, v. 1, n. 2, p. 183–187, 1986.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. **Biochimica et Biophysica Acta (BBA)-Protein Structure**, v. 405, n. 2, p. 442–451, 1975.

MENDOLIA, I. et al. Convolutional architectures for virtual screening. **BMC Bioinformatics**, v. 21, 16 set. 2020.

MESTRES, J. Representativity of target families in the Protein Data Bank: impact for family-directed structure-based drug discovery. **Drug discovery today**, v. 10, n. 23–24, p. 1629–1637, 2005.

MITCHELL B.O., J. B. O. Machine learning methods in chemoinformatics**. Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 4, n. 5, 2014.

MSONTHQ, J. D. 164. Uncinatone, a New Antifungal Hydroquinone Diterpenoid from. **Helvetica Chimica Acta**, v. 68, n. 1985, 2000.

NDJOUBI, K. O. et al. Antimycobacterial, cytotoxic, and antioxidant activities of abietane diterpenoids isolated from *Plectranthus madagascariensis*. **Plants**, v. 10, n. 1, p. 1–12, 1 jan. 2021.

PÉREZ-SIANES, J.; PÉREZ-SÁNCHEZ, H.; DÍAZ, F. Virtual Screening Meets Deep Learning. **Current Computer-Aided Drug Design**, v. 15, n. 1, p. 6–28, 19 out. 2018a.

PÉREZ-SIANES, J.; PÉREZ-SÁNCHEZ, H.; DÍAZ, F. Virtual Screening Meets Deep Learning. **Current Computer-Aided Drug Design**, v. 15, n. 1, p. 6–28, 19 out. 2018b.

PICKING, D. et al. *Hyptis verticillata* Jacq: A review of its traditional uses, phytochemistry, pharmacology and toxicology. **Journal of Ethnopharmacology**, v. 147, n. 1, p. 16–41, 2013.

RADULOVIĆ, N.; DENIĆ, M.; STOJANOVIĆ-RADIĆ, Z. Antimicrobial phenolic abietane diterpene from *Lycopus europaeus* L. (Lamiaceae). **Bioorganic & medicinal chemistry letters**, v. 20, n. 17, p. 4988–4991, 2010.

RODRIGUES, R. P. et al. Virtual screening strategies in drug design. **Revista Virtual de Quimica**, v. 4, n. 6, p. 739–776, 2012.

RODRÍGUEZ, B. et al. Neo-clerodane insect antifeedants from *Scutellaria galericulata*. **Phytochemistry**, v. 33, n. 2, p. 309–315, 1993.

ROLIM, T. L. et al. Toxicity and antitumor potential of *Mesosphaerum sidifolium* (Lamiaceae) oil and fenchone, its major component. **BMC Complementary and Alternative Medicine**, v. 17, n. 1, p. 347, 3 dez. 2017.

SÁ- FILHO, G. F. DE et al. Plantas medicinais utilizadas na caatinga brasileira e o potencial terapêutico dos metabólitos secundários: uma revisão. **Research, Society and Development**, v. 10, n. 13, p. e140101321096, 9 out. 2021.

SCHNEIDER, G.; DOWNS, G. **Machine learning methods in QSAR modelling.** QSAR & Combinatorial Science, v. 22, n. 5, p. 485-486, 2003.

SCOTTI, L. et al. *In-silico* Analyses of Natural Products on *Leishmania* Enzyme Targets. **Mini-Reviews in Medicinal Chemistry**, v. 15, n. 3, p. 253–269, 12 mar. 2015.

SCOTTI, M. et al. SistematX, an Online Web-Based Cheminformatics Tool for Data Management of Secondary Metabolites. **Molecules**, v. 23, n. 1, p. 103, 3 jan. 2018.

SHAN, W. et al. Convolutional Neural Network-based Virtual Screening. **Current Medicinal Chemistry**, v. 28, n. 10, p. 2033–2047, 26 maio 2020.

SHANG, X. et al. The genus Scutellaria an ethnopharmacological and phytochemical review. **Journal of Ethnopharmacology**, v. 128, n. 2, p. 279–313, mar. 2010.

SHARMA, D. et al. In Vitro Efficacy of Lipid Conjugated Peptidomimetics Against *Mycobacterium smegmatis*. **International Journal of Peptide Research and Therapeutics**, v. 26, n. 1, p. 531–537, 1 mar. 2020.

SILVA, A. B. 5'-Deacetoxi-hiptenolideo, uma α-pirona isolada de *Hyptis Macrostachys* Benth. **Monografia**, p. 1-54, Universidade Federal da Paraíba, João Pessoa: 2014. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/593>. Acesso em: 27 nov. 2021.

SILVA, A. B. ESTUDO FITOQUÍMICO DE *Leptohyptis macrostachys* (BENTH.) HARLEY & J.F.B.PASTORE E *Mesosphaerum sidifolium* (L'HÉRIT.) HARLEY & J.F.B.PASTORE DA SUBTRIBO HYPTIDINAE (LAMIACEAE). Dissertação, p. 1-139, Universidade Federal da Paraíba, João Pessoa, 2016. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/20694>. Acesso em: 24 nov. 2021.

Simões, C. M. O., Schenkel, E. P., de Mello, J. C. P., Mentz, L. A., & Petrovick, P. R. **Farmacognosia: do produto natural ao medicamento**. Artmed Editora, 2016.

SIMÕES, R. S. et al. Transfer and multi-task learning in QSAR modeling: advances and challenges. **Frontiers in pharmacology**, v. 9, p. 74, 2018.

SOARES, A. DE S.; PASTORE, J. F. B.; JARDIM, J. G. Lamiaceae no Rio Grande do Norte, Brasil. **Rodriguésia**, v. 70, 2019.

TAMBURINI, B. et al. Role of hematopoietic cells in *Mycobacterium tuberculosis* infection. **Tuberculosis**, v. 130, p. 102109, 2021.

TODESCHINI, R.; CONSONNI, V. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references. **John Wiley & Sons**, v. 41, 2009.

URITU, C. M. et al. Medicinal Plants of the Family Lamiaceae in Pain Therapy: A Review. **Pain Research and Management**, v. 2018, p. 1–44, 2018.

VARELA-RIAL, A.; MAJEWSKI, M.; DE FABRITIIS, G. Structure based virtual screening: Fast and slow. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, p. 1544, 2021.

VARNEK, A. Tutorials in chemoinformatics. [s.l.] **John Wiley & Sons**, 2017.

VERSION 7.0. **Kode srl, Dragon (software for molecular descriptor calculation)**. Disponível em: <https://chm.kode-solutions.net>.

VIEIRA DE BARROS, A. et al. AS RIQUEZAS DA CAATINGA E SEU POTENCIAL FARMACOLÓGICO: UMA REVISÃO SISTEMÁTICA THE RICHES OF CAATINGA AND ITS PHARMACOLOGICAL POTENTIAL: A SYSTEMATIC REVIEW. **South American Journal**, v. 8, n. 1, p. 771–791, 2021.

VUK, M.; CURK, T. ROC Curve, Lift Chart and Calibration Plot Metodološki zvezki. P. 89-108, 2006.

WAN, J. et al. Ent-abietanoids isolated from *Isodon serra*. **Molecules**, v. 22, n. 2, 1 fev. 2017.

YAO, J.-L. et al. A Review on the Terpenes from Genus *Vitex*. **Molecules**, v. 21, n. 9, p. 1179, 6 set. 2016.

ZHAO, F. et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. **BMC Biology**, v. 19, n. 1, 1 dez. 2021.

Capítulo 2

Neste capítulo, foi construído um banco de dados de moléculas de diterpenos isolados da família Lamiaceae, a partir de uma revisão bibliográfica que foi realizada utilizando busca eletrônica nas bases de dados SciFinder (https://scifinder.cas.org/) e Web of Science (https://clarivate.com/products/web-of-science/), abrangendo diversos artigos publicados entre os anos de 1980 a 2017. Foram adicionados ao SistematX (http://sistematx.ufpb.br) 4115 estruturas químicas diferentes que correspondem a 6386 ocorrências botânicas, que estão distribuídas em 8 subfamílias, 66 gêneros, 639 espécies diferentes e 4880 localizações geográficas.

Foi realizado uma análise quimiotaxonômica entre as subfamílias de Lamiaceae, empregando os descritores moleculares e redes neurais não supervisionadas. Para tanto, foram gerados Mapas Auto Organizáveis (SOMs) utilizando descritores moleculares dos diterpenos e suas respectivas ocorrências botânicas. Em todos os mapas obtidos, verificou-se uma taxa de acerto superior a 80%, evidenciando uma separação das subfamílias de Lamiaceae, corroborando com os dados morfológicos e moleculares propostos por Li.

Sendo assim, através deste estudo quimiotaxonômico foi possível predizer a localização de um diterpeno em uma determinada subfamília e auxiliar na prospecção de metabólitos secundários com características estruturais específicas, como compostos com potencial atividade biológica.

Computer-Aided Chemotaxonomy and Bioprospecting Study of Diterpenes of the Lamiaceae Family

Andreza Barbosa Silva Cavalcanti, Renata Priscila Costa Barros, Vicente Carlos de Oliveira Costa, Marcelo Sobral da Silva, Josean Fechine Tavares, Luciana Scotti and Marcus Tullius Scotti*

Post-Graduate Program in Natural Synthetic Bioactive Products, Federal University of Paraiba, João Pessoa58051-900, Paraíba, Brazil; andreza.jp.pb@gmail.com (A.B.S.C.); renatapcbarros@gmail.com (R.P.C.B.); vicente@ltf.ufpb.br (V.C.d.O.C.); marcelosobral.ufpb@gmail.com (M.S.d.S.); josean@ltf.ufpb.br (J.F.T.); luciana.scotti@gmail.com (L.S.)

***** Correspondence: mtscotti@gmail.com; Tel.: +55-83-998690415

Received: 4 October 2019; Accepted: 27 October 2019; Published: 30 October 2019

Artigo publicado na revista: **Molecules**, 2019, 24, 3908.

DOI: 10.3390/molecules24213908

Fator de impacto: 4.411

Abstract: Lamiaceae is one of the largest families of angiosperms and is classified into 12 subfamilies that are composed of 295 genera and 7775 species. It presents a variety of secondary metabolites such as diterpenes that are commonly found in their species, and some of them are known to be chemotaxonomic markers. The aim of this work was to construct a database of diterpenes and to use it to perform a chemotaxonomic analysis among the subfamilies of Lamiaceae, using molecular descriptors and self-organizing maps (SOMs). The 4115 different diterpenes corresponding to 6386 botanical occurrences, which are distributed in eight subfamilies, 66 genera, 639 different species and 4880 geographical locations, were added to SistematX. Molecular descriptors of diterpenes and their respective botanical occurrences were used to generate the SOMs. In all obtained maps, a match rate higher than 80% was observed, demonstrating a separation of the Lamiaceae subfamilies, corroborating with the morphological and molecular data proposed by Li et al. Therefore, through this chemotaxonomic study, we can predict the localization of a diterpene in a subfamily and assist in the search for secondary metabolites with specific structural characteristics, such as compounds with potential biological activity.

Keywords: Lamiaceae; database; diterpenes; chemotaxonomic; SOMs

1. Introduction

Historically, natural products have been used as sources to treat, cure and prevent diseases [1]. The greatest contribution of these natural products occurs through plants, which can be classified according to their chemical constitution, and this classification is defined as chemotaxonomy. A wide variety of studies include the chemotaxonomic classification of secondary metabolites; among the most investigated compounds are phenolics, alkaloids, terpenoids and nonprotein amino acids [2].

Lamiaceae is one of the largest families of Angiosperms, the largest family of Lamiales, an order comprising 26 families and more than 20,000 species [3]. This family is classified into 12 subfamilies, which are composed of 295 genera and 7775 species [4]. Their species are usually represented by herbs and shrubs that are distributed throughout the world in tropical and temperate regions [3,5].

This family accumulates substances with very diverse structures and many of them are reported as chemotaxonomic markers at all levels: subfamily, genus and species

[6]. The main secondary metabolites isolated from Lamiaceae species are monoterpenes [7], sesquiterpenes [8], diterpenes [9], triterpenes [10], pyrones [11], iridoids [12], phenolic compounds [13] and flavonoids [14]. Among these, diterpenes are more prominent as chemotaxonomic markers because they are easily found in most species of this family [6,15–20].

Several botanical studies have demonstrated the classification of Lamiaceae in the level of subfamilies. In the work of Harley et al. [21], it was observed that 236 genera are distributed in seven subfamilies: Ajugoideae, Lamioideae, Nepetoideae, Prostantheroideae, Scutellarioideae, Symphorematoideae and Viticoideae, although 10 genera were left unassigned at the subfamily level.

Recently, Li et al. [3] presented a review containing several findings that strengthen arguments for a new classification of the family Lamiaceae, reporting for each subfamily its phylogenetics and morphology. They observed through DNA analysis the presence of three new subfamilies, making up a total of 10 subfamilies (Figure 1). Of the ten genera that were unclassified in the study by Harley et al. [21], the only two that were not allocated to a subfamily were *Callicarpa* and *Tectona*. However, these two genera are inserted between the subfamilies in positions that corroborate with the phylogeny (Figure 1).

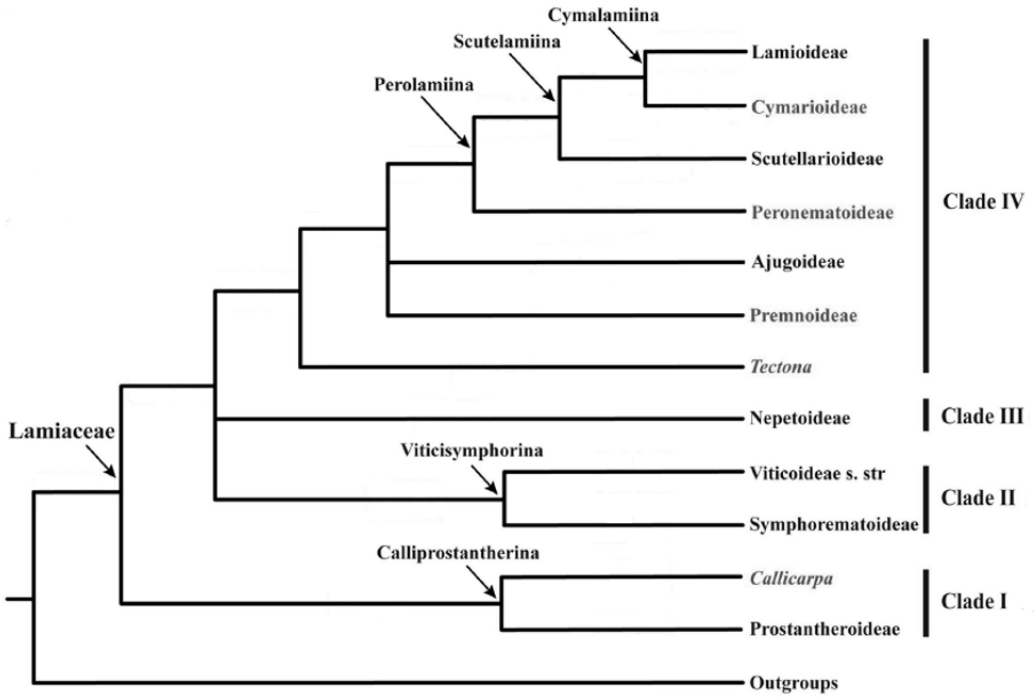


Figure 1. Phylogenetic diagram of Lamiaceae subfamilies (adapted from Li et al. [3]).

In the study by Li et al. [22], the presence of two new subfamilies, Callicarpoideae and Tectonoideae, was determined, which in the previous study had not been classified. Therefore, the current Lamiaceae classification is composed of 12 subfamilies arranged in four clades, thus facilitating the organization of genera and species [3,22].

The biological and physicochemical properties of the molecules can be predicted through molecular descriptors, which are the result of the conversion of the symbolic representation of a chemical structure into a useful number [23], and there are several software packages to generate molecular descriptors, such as Dragon 7.0 [24]. The descriptor can be used to obtain chemical patterns that, to be visualized, need the creation of computational models that can be obtained using several algorithms such as the use of artificial neural networks (ANNs).

ANNs are defined as a mathematical model inspired by the neural structure of intelligent organisms, in which several nodes, called neurons, are interconnected in a network-like structure [25,26]. In the process of identifying and classifying patterns, the commonly used ANNs' unsupervised architecture is the self-organizing map (SOM). This is an unsupervised method capable of providing multivariate data maps in a two-dimensional (2-D) grid. It results in the clustering of similar patterns next to each other and has been used successfully in different studies that use database chemistry, including chemotaxonomic studies [25–31].

In the search for secondary metabolite banks already isolated from the Lamiaceae family, we can use databases that provide information about the compounds, such as biological, biogeographical and taxonomic data [32]. Some of these tools are commercially available or freely available, such as the Bioassay Nucleus, Biosynthesis and Ecophysiology of Natural Products (NuBBE) [33], Dictionary of Natural Products (DNP) [34], NAPRALERT [35] and Marinlit for natural marine products [36].

SistematX has a different relationship to the other databases available on the web, in that it is possible to use a browser to directly add and manage the data useful to the academic community about the secondary metabolites, such as research by chemical structure, SMILES code, compound names as well as information-specific species for taxonomic classification (from family to species) and the geographic location of the species from which the compounds were isolated [32].

Thus, the aim of this work is to construct a database of diterpenes from the Lamiaceae family and extract information for chemotaxonomic analysis among the subfamilies, using the molecular descriptors and SOMs, and comparing the results with the phylogenetic classification proposed by Li et al. [3]. This will test if it is possible to predict the botanical occurrence in its corresponding subfamily.

2. Results and Discussion

2.1 Database

The database is composed of diterpenes isolated from species of the family Lamiaceae; it comprises 4115 different chemical structures and corresponds to 6386 botanical occurrences and 4880 geographical locations. The number of occurrences for a superior taxon is defined counting how many times a compound appears in determined species belonging to that taxon. All data are available in the SistematX tool (https://sistematx.ufpb.br). As shown in Table 1, the 4115 diterpene molecules are distributed in eight subfamilies, 66 genera and 639 different species of the Lamiaceae family. The subfamily Nepetoideae presents the greatest number of genera, species and botanical occurrences. Of the total number of botanical occurrences, only seven botanical occurrences were unclassified at subfamily level, therefore totaling 6379.

Table 1. Lamiaceae subfamilies listed according to Li et al. [3]. Abbreviations, botanical data, number of diterpenes and chemical occurrences added and used in SistematX (https://sistematx.ufpb.br).

Subfamily	Acronym	Genera	Species	Diterpenes	Occurrences
Ajugoideae	Aju	7	99	580	856
Callicarpoideae	Cal	1	14	71	86
Lamioideae	Lam	23	188	601	1183
Nepetoideae	Nep	30	289	2433	3644
Peronematoideae	Per	1	1	7	7
Premnoideae	Pre	2	7	85	92
Scutellarioideae	Scu	1	31	286	342
Viticoideae	Vit	1	10	130	169
Total		66	639	4115	6379

2.2 Self-Organizing Maps and Molecular Descriptors Applied in the Chemotaxonomy of Lamiaceae Subfamilies

From the botanical occurrences of the diterpenes obtained from the Lamiaceae family, 108 molecular descriptors were generated for each molecular structure using

Dragon 7.0 software [24]. The botanical occurrences were classified into four subfamilies and the values of the descriptors were used as input data for the SOM Toolbox 2.0 software [37]. The subfamilies selected for analysis were those that presented the highest number of botanical occurrences making possible the pattern recognition of the distribution of diterpenes in Lamiaceae (Table 1). Then, the self-organizing matrix for each molecule was calculated, dividing the samples into groups according to the similarity and after comparing the SOM with the classification proposed by Li et al. [3].

In the maps depicted, the chemical occurrences of certain subfamilies occupy regions that are labeled by the following colors:

1. Clade III (Nepetoideae), red;
2. Clade IV (Ajugoideae, Lamioideae and Scutellarioideae), lilac;
3. Ajugoideae, blue;
4. Lamioideae, green;
5. Scutellarioideae, dark blue.

The SOM that was obtained using the occurrences of the diterpenes of clade III (Nep) and clade IV (Aju, Lam and Scu) subfamilies showed a total hit rate of 86.3%, with 6025 occurrences and 5200 hits (Table 2). The SOM generated using fingerprint to analyze the correspondence of botanical occurrences of clade III and clade IV subfamilies resulted in a total hit rate of 89.5%. These data corroborate a good separation of the subfamilies because even though different descriptors were used, the results were similar (Table 2).

Table 2. Results of the self-organizing map with the values of the occurrences and the number of correct hits for clade III (Nep) and clade IV (Aju, Lam and Scu) of the family Lamiaceae, using the descriptors generated by the program DRAGON 7.0 [24].

			Molecular Descriptors		Fingerprint	
Subfamily	**Diterpenes**	**Occurrences**	**No. of Hits**	**% of Hits**	**No. of Hits**	**% of Hits**
Clade III	2433	3644	3252	89.2	3366	92.3
Clade IV	1453	2381	1948	81.8	2031	85.3
Total	3886	6025	5200	86.3	5397	89.5

The SOM (Figure 2) shows a clear separation between the botanical occurrences of clade III (red) and clade IV (lilac), reaffirming the phylogenetic analysis performed by Li et al. (Figure 1) [3]. Analyzing the SOM, there is a chemical pattern that shows a region in which the subfamily Nep (red) occupies many neurons distributed by the map,

being the one with the highest number of occurrences (3644) and the best rate of success 89.2% (Table 2). The predictive performance of the SOM for the five training and test sets that were generated from the original set can be visualized in Table 3. The applicability domain (AD) was reliable for more than 99% of the predictions of the test set. The average match rate for the five test sets (85.4%) is very close to that of the training (86.4%). The clade III (Nep subfamily) shows the highest match rate values for training sets (88.6%) and tests (88.3%), while clade IV (subfamilies Aju, Lam and Scu) showed 82.1% and 81% for training and test sets, respectively.

Table 3. Summary of results of training and test match (%) of clade III (Nep) and clade IV (Aju + Lam + Scu).

Subfamily	Train Set 1	Train Set 2	Train Set 3	Train Set 4	Train Set 5	Average
Clade III	91.4	87.1	89.5	88.5	89.4	88.6
Clade IV	79.9	84.8	80.1	83.2	82.3	82.1
Total	86.9	86.2	85.8	86.4	86.6	86.4
Subfamily	**Test Set 1**	**Test Set 2**	**Test Set 3**	**Test Set 4**	**Test Set 5**	**Average**
Clade III	89.4	87.7	89.8	87.5	87.2	88.3
Clade IV	76.9	84.2	84.2	78.6	81.1	81.0
Total	84.5	86.3	87.6	84.0	84.8	85.4

Chemotaxonomy analysis was also performed using other machine learning algorithms: support vector machine (SVM), which is a supervised machine learning algorithm, and k- nearest neighbors (k-NN), which is an instance-based algorithm. Results are shown in Table 4 for the analysis performed on the SOM by clade. It can be observed that, as in the SOM, the models generated with SVM and k-NN obtained very similar results and with high hit rates.

Table 4. Summary of test match (%) corresponding to the results obtained from 5-fold models using self-organizing map (SOM), support vector machine (SVM) and k-nearest neighbors (k-NN) algorithms for clade III (Nep) and clade IV (Aju + Lam + Scu).

Subfamily	SOM Average	SOM fingerprint Average	SVM Average	k-NN Average
Clade III	88.3	88.1	92.2	96.2
Clade IV	81.0	80.0	88.9	92.2
Total	85.4	89.5	90.9	94.6

The applicability domain (AD) was reliable for over 99% of the test set predictions for all algorithms used: SOM with molecular descriptor, SOM with fingerprint, SVM and k-NN.

The most significant descriptors for the clustering the diterpenes of the Ajugoideae, Lamioideae, Scutellarioideae (clade IV) and Nepetoideae (clade III) subfamilies are also shown in Figure 2. The U-matrix shows the distances between the neighboring map unit, where high values indicate a border of a cluster and uniform areas of low values indicate the clusters themselves (Figure 2a). The subfamily of clade III shows a high value for the following descriptors, which are shown in black in Figure 2a: atoms-centered descriptor O-056 that encodes alcohol and functional group count nArOH that encodes the number of aromatic hydroxyls. The diterpenes of the clade IV subfamilies present high values for the ring descriptor NRS that encodes the number of ring systems (Figure 2a).

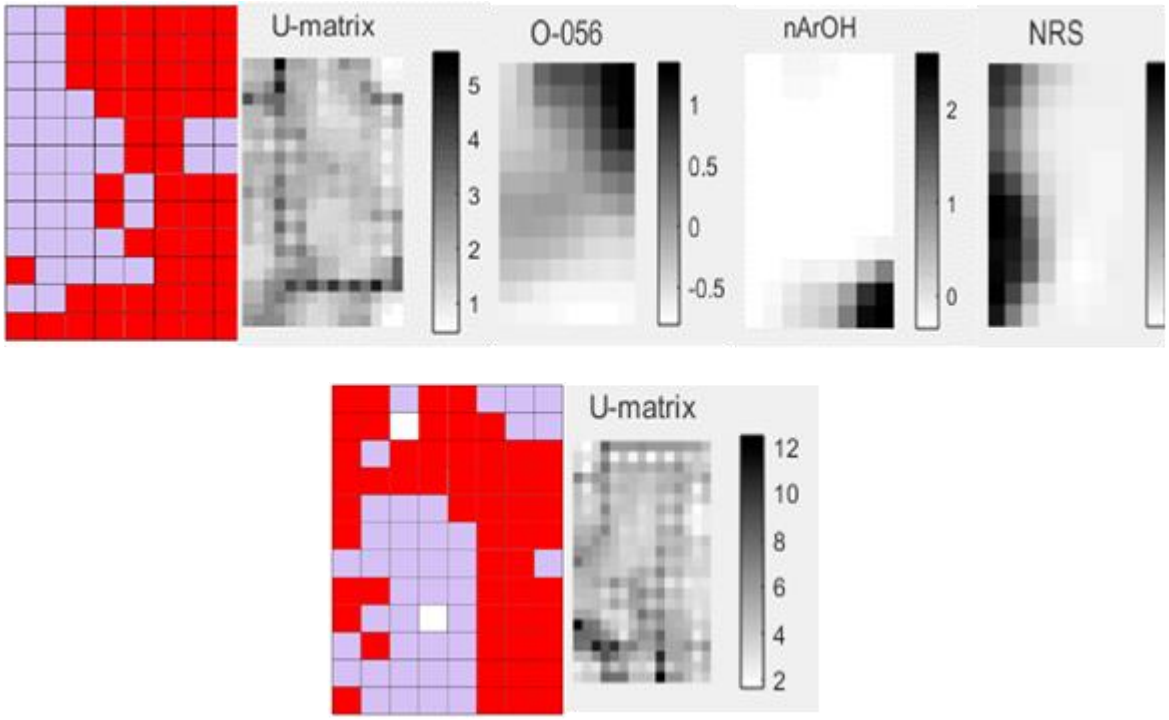


Figure 2. Self-organizing map obtained by classification of the subfamilies of clade III (red) and clade IV (lilac) and generated descriptors: (a) SOM → molecular descriptors; U-matrix; O-056, nArOH and NRS. (b) SOM → fingerprint and U-matrix.

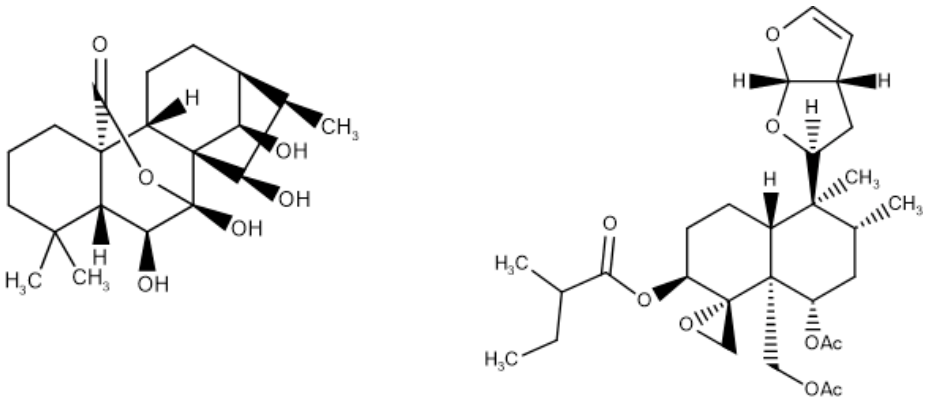
In analyzing the individual descriptors, it was verified in the descriptor of atom-centered fragments, O-056 (alcohol), that its highest value was attributed to diterpene 1 (Figure 3) due to the presence of four alcohols. This diterpene is popularly known as isorosthin J [38,39] and belongs to the subfamily Nepetoideae (clade III). The diterpene 2 (Figure 3), known as ajubractin A [40], belongs to the subfamily Ajugoideae (clade IV) and presents the null value for the descriptor O-056. It was observed that diterpene 3 (Figure 3), known as plectranthol A [41], has the highest value of the nArOH

descriptor, with the presence of four aromatic hydroxyls, whereas the lowest value, null, for this descriptor was attributed to diterpene 4, lupulin A [42–45] (Figure 3).

It was reported in the literature that plectranthol A (3) shows antioxidant activity [41] and, according to this chemotaxonomic study, it is observed that it can be found in a species belonging to the subfamily Nepetoideae of clade III (red) (Figure 2a), whereas lupulin A has potential antibacterial activity [42] being commonly found in species of clade IV subfamilies, Ajugoideae and Scutellarioideae [42–45] (Figures 2 and 3).

By examining the NRS descriptor (Figure 2a), it was found that diterpene 5 (Figure 3), which is known as scutalpin L [46,47], presented the highest value for this descriptor, having in its molecule four ring systems, occurring in the subfamily Scutellarioideae of clade IV. Diterpene 6 (crassifol) [48] of the subfamily Nepetoideae shows a null value for the NRS descriptor because it has an acyclic structure (Figure 3).

This confirms that there is a chemical profile of diterpenes, which shows that the subfamilies of clade IV present diterpenes with more ring systems and that the subfamily Nepetoideae (clade III) has molecules rich in hydroxyl groups attached to aromatic and nonaromatic groups.



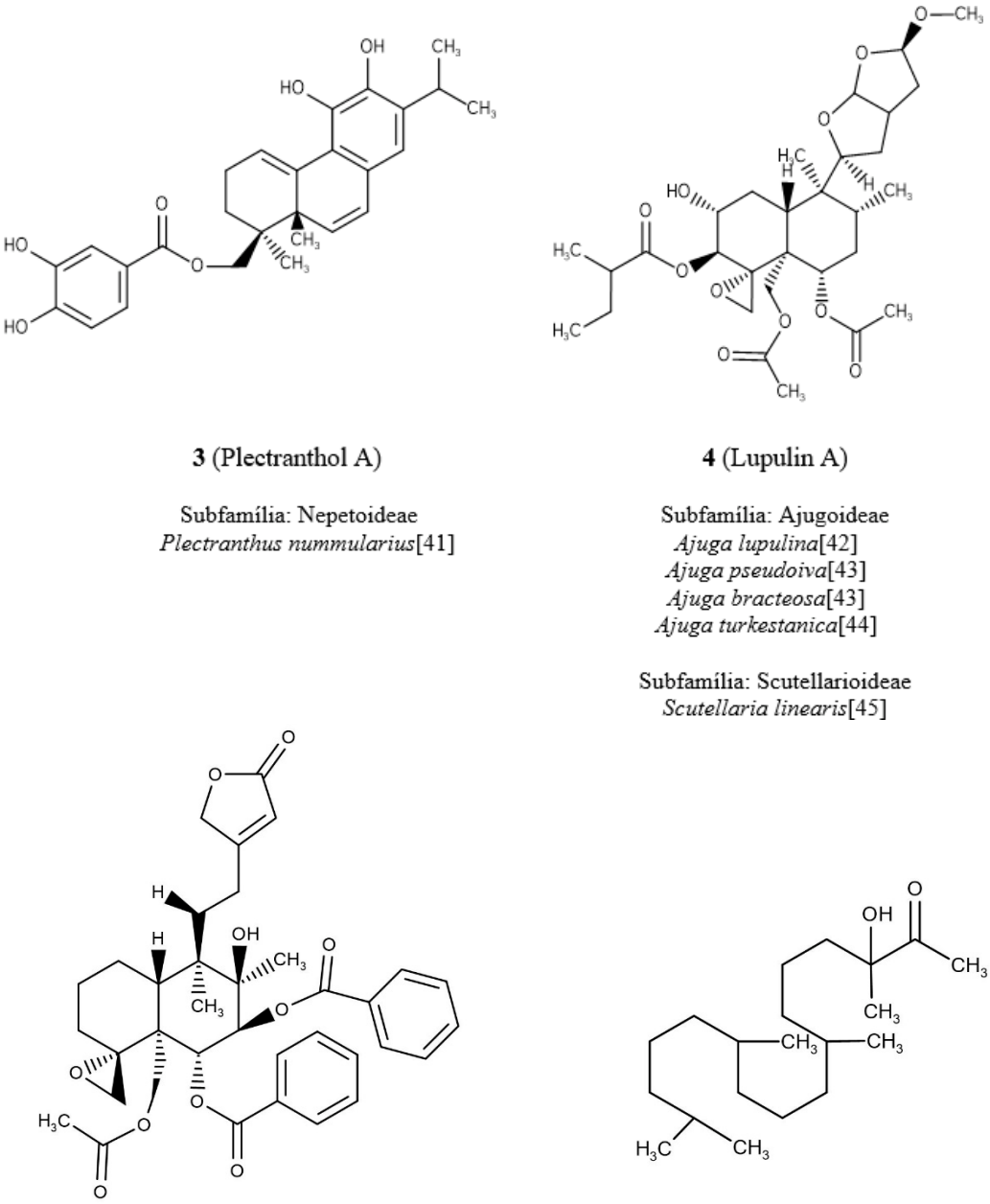
1 (Isorosthin J)

Subfamília: Nepetoideae
Isodon rosthornii[38,39]

2 (Ajubractin A)

Subfamília: Ajugoideae
Ajuga bracteosa[40]

Figure 3.
Cont.



3 (Plectranthol A)

Subfamília: Nepetoideae
Plectranthus nummularius[41]

4 (Lupulin A)

Subfamília: Ajugoideae
Ajuga lupulina[42]
Ajuga pseudoiva[43]
Ajuga bracteosa[43]
Ajuga turkestanica[44]

Subfamília: Scutellarioideae
Scutellaria linearis[45]

5 (Scutalpin L)

Subfamily: Scutellarioideae
Scutellaria alpina [46,47]

6 (Crassifol)

Subfamily: Nepetoideae
Nepeta crassifolia [48]

Figure 3. Chemical structures of the diterpenes located in the SOM of clade III (Nep) and clade IV (Aju, Lam, Scu) and their respective botanical occurrences.

The SOM generated to analyze the correspondences of the 2381 diterpene botanical occurrencesof the clade IV subfamilies (Aju, Lam and Scu) resulted in a total hit rate of 91.4% (Table 5). It isalso observed that the subfamily Lam presents

the best hit rate with 94.8% and the largest number of occurrences and compounds of clade IV; its structural diversity in terms of diterpenes is shown in the SOM (Figure 4). The subfamily Scu shows a hit rate of 81.3%, revealing a clear separation of these subfamilies because all the subfamilies present an accuracy greater than 80%.

Table 5. Results of the self-organizing maps with the occurrence values and the number of correct hits for the subfamilies belonging to clade IV (Aju, Lam and Scu), using the descriptors generated by the Dragon 7.0 program.

			Molecular Descriptors		Fingerprint	
Subfamily	Diterpenes	Occurrences	No. of Hits	% of Hits	No. of Hits	% of Hits
Aju	580	856	776	90.7	742	86.6
Lam	601	1183	1122	94.8	1122	94.8
Scu	286	342	278	81.3	320	93.5
Total	1467	2381	2176	91.4	2184	91.7

Using fingerprint, rates of accuracy were observed close to those obtained using the molecular descriptors; the subfamily Lam had the same hit rate 94.8% in the fingerprint (Table 5). This information supports a good SOM rating performance even when using two different types of descriptors.

Table 6 shows a significant correspondence in the training and test sets of the Aju, Lam and Scu subfamilies. Once more, the AD was reliable for more than 99% of the predictions of the test set. Lamioideae have higher match values: 95.9 and 94.1% for the training and testing, respectively. Scutellarioideae shows lower matching values in the training models with a mean of 76.2% and similar performance in the test results (68.1%). All the total training and test results show a level of significance higher than 60%.

Table 6. Summary of the results of training and test match (%) of Aju, Lam and Scu.

Subfamily	Train Set 1	Train Set 2	Train Set 3	Train Set 4	Train Set 5	Average
Aju	86.6	90.2	92.8	88.6	87.6	89.2
Lam	96.4	96.2	95.7	94.6	96.5	95.9
Scu	78.4	75.9	67.2	76.6	82.8	76.2
Total	91.4	91.1	90.6	89.9	91.3	90.9
Subfamily	**Test Set 1**	**Test Set 2**	**Test Set 3**	**Test Set 4**	**Test Set 5**	**Average**
Aju	93.0	90.1	89.5	90.6	82.5	89.1
Lam	93.6	91.1	96.6	95.8	93.2	94.1

Scu	63.8	67.6	54.4	70.6	84.1	68.1
Total	89.1	87.4	88.0	90.3	88.0	88.6

Chemotaxonomy analysis was also performed using other machine learning algorithms, i.e., support vector machine (SVM), which is a supervised machine learning algorithm, and k- nearest neighbors (k-NN), which is an instance-based algorithm. The results are shown in Table 7 for the analysis performed on the SOM by subfamilies belonging to clade IV. It can be observed that, as in the SOM, the models generated with SVM and k-NN obtained very similar results, with high hit rates. The applicability domain (AD) was reliable for over 99% of the test set predictions for all algorithms used: SOM with molecular descriptor, SOM with fingerprint, SVM and k-NN.

Table 7. Summary of test match (%) corresponding of the results obtained 5-fold models using SOM, SVM and k-NN algorithm of the subfamily Aju, Lam and Scu.

Subfamily	SOM Average	SOM Fingerprint Average	SVM Average	k-NN Average
Aju	89.1	83.6	94.6	95.3
Lam	94.1	87.3	97.0	97.6
Scu	68.1	95.8	87.4	88.0
Total	88.6	92.7	94.7	95.4

In analyzing the SOM and descriptors obtained only from clade IV, the diterpenes of the Ajugoideae, Lamioideae and Scutellarioideae subfamilies that make up this clade were used (Figure 4a). In the map, we can see that there is a proximity between Lam (green) and Aju (light blue), as well as Aju (light blue) with Scu (dark blue), therefore, the pattern of the botanical occurrence of diterpenes does not corroborate with the phylogenetic classification proposed by Li et al. [3], who report that Lam (green) would be closer to Scu (dark blue) than Aju (light blue).

As shown in Figure 4, the self-organizing map obtained by fingerprint showed similarity in the separation of diterpenes when compared to the map obtained by the fragment descriptors.

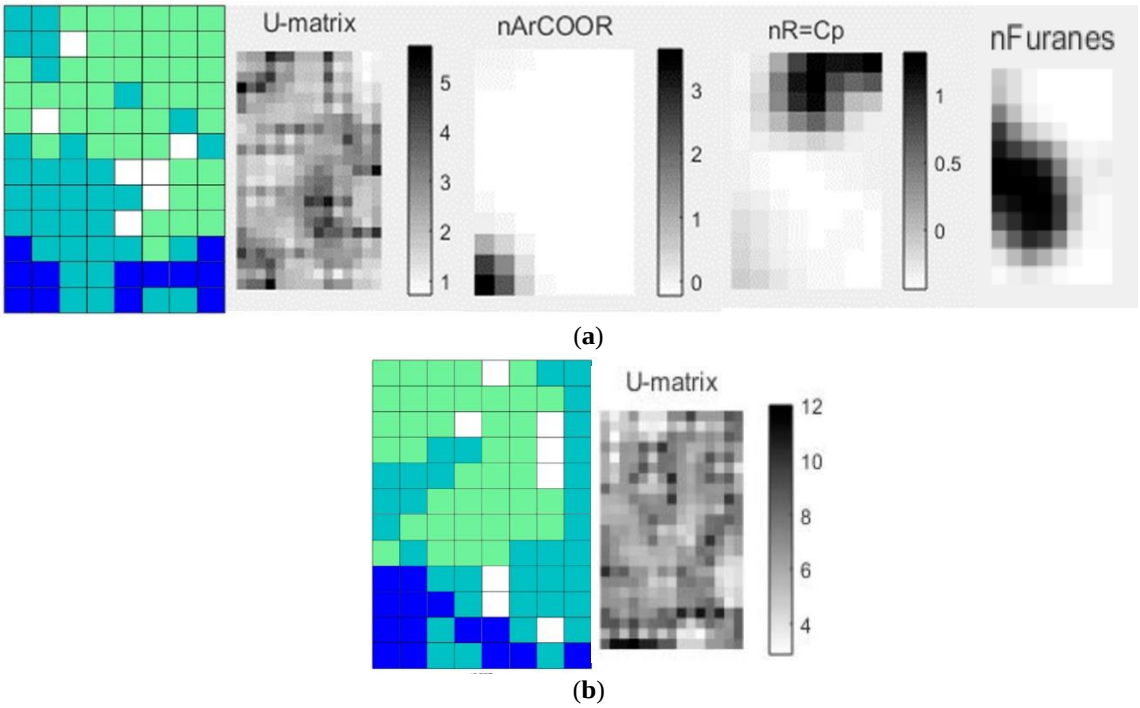


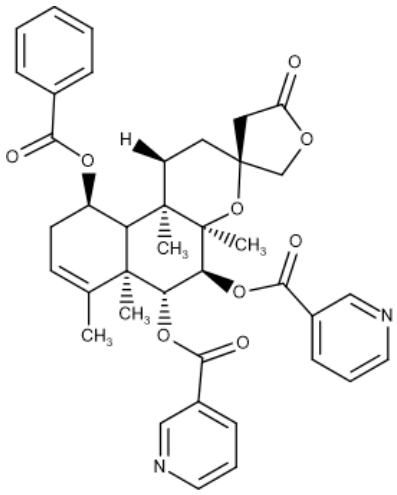
Figure 4. Self-organizing map obtained by the classification of the subfamilies Aju (light blue), Lam (green), Scu (dark blue) and generated descriptors: (a) SOM → molecular descriptors; U-matrix; nArCOOR; nR = Cp and nFuranes. (b) SOM → fingerprint and U-matrix.

Analyzing the descriptors shown in Figure 4a, in the black color for higher values, one realizes that the diterpenes of the Scu subfamily display a high value for the nArCOOR (number of aromatic esters) descriptor; secondary metabolites of subfamily Lam show high values in the descriptor nR = Cp (number of primary C terminals—sp2) and the subfamily Aju has molecular structures with higher values of the descriptor nFuranes (number of furans).

The diterpene 7 (Figure 5) shows the highest value for the nArCOOR descriptor because in its structure it has three aromatic esters. It is commonly known as scutebatin B [49], being found in the subfamily Scutellarioideae (dark blue) (Figure 4a), and the study of its isolation verified its inhibitory effects on the production of nitric oxide aromatic esters induced by lipopolysaccharide in macrophages [49]. We can observe in the descriptor nArCOOR that the white spaces are formed by regions of smaller values, being related to the diterpenes of Lamioideae (green) and Ajugoideae (light blue) (Figure 4a). Thus, we have as example diterpene 8 (Figure 5), known as cyllenin A [50,51], which does not have aromatic ester groups and belongs to the subfamily Lamioideae.

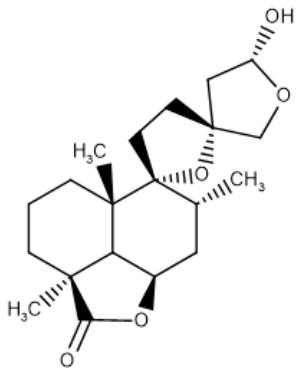
We investigated the highest value reported in the descriptor nR = Cp, which was attributed to diterpene 9 (Figure 5) which is known as sclarene [7]; with three sp2 terminal carbons, this diterpene occurs in the subfamily Lamioideae (green) (Figure 4a). The lowest value of the descriptor nR = Cp corresponds to the diterpene 10 (Figure 5), which does not present any terminal carbon sp2 and is located in the subfamily Ajugoideae (light blue). Diterpene 10 is known as ajugamarin A1 [43] and shows a potential neuroprotective effect [52].

The diterpene 11 (Figure 5), teubrevin G [53,54], presents the highest value for the nFurane descriptor because there are two furan rings. Observing the descriptor in the black region, which represents higher values, and comparing with the map matches with the same region in which the diterpenes of Ajugoideae occupy confirms that this diterpene occurs in the subfamily Ajugoideae. The diterpene 12 (Figure 5), known as sidendrodiol [7,55–57], belongs to the species that occur in the subfamily Lamioideae and does not have furan groups.



7 (Scutebatin B)

Subfamília: Scutellarioideae
Scutellaria barbata[49]



8 (Cyllenin A)

Subfamília: Lamioideae
Marrubium peregrinum[50]
Marrubium globosum[51]

Figure 5.
Cont.

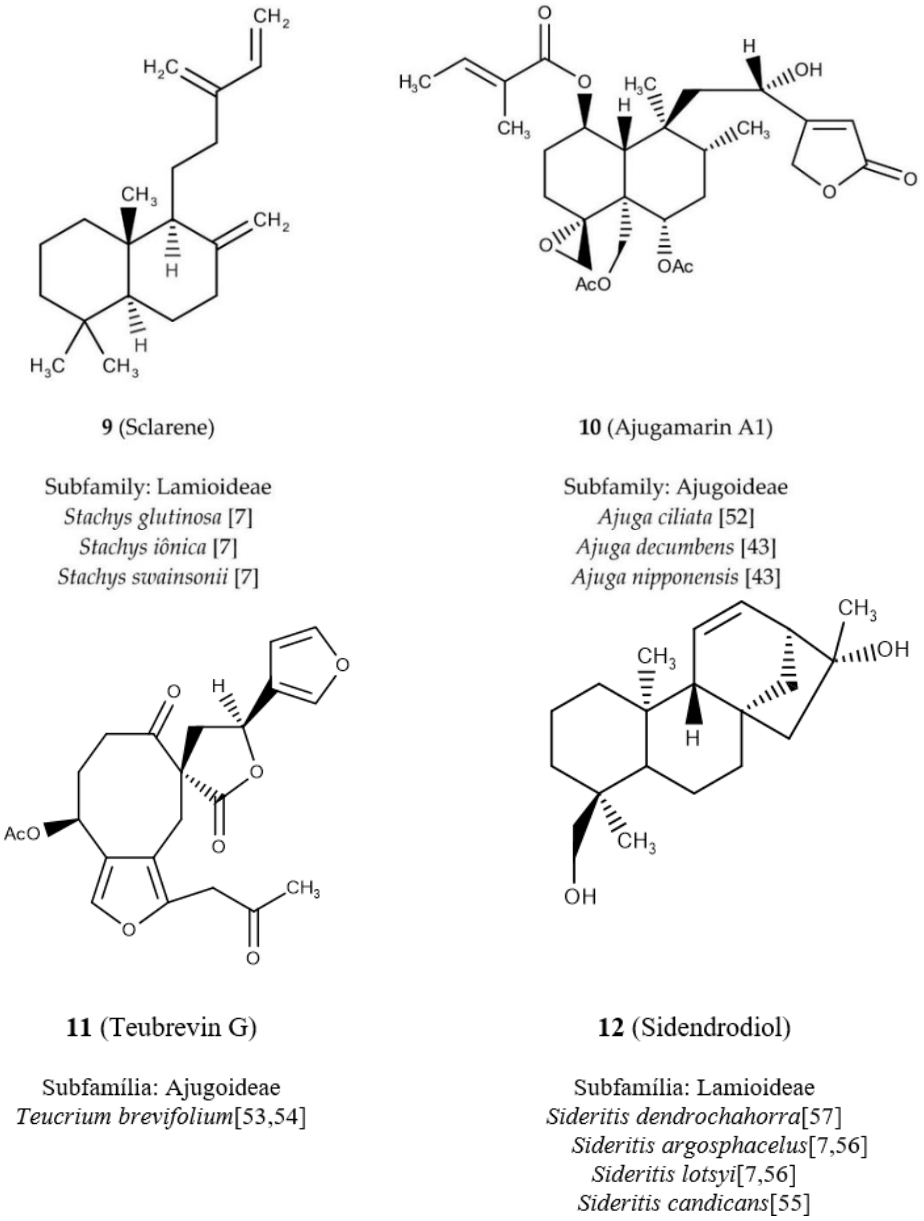


Figure 5. Chemical structures of diterpenes located in SOM and their respective botanical occurrences.

The Lamiaceae family includes the genus Scutellaria, which belongs to the subfamily Scutellarioideae, and has a cosmopolitan distribution of around 360 species worldwide and in different climatic regions. A majority of its growing species in Asia have a long tradition in Chinese folk medicine [46]. Several studies indicate that diterpenes are commonly found in these species. Isodon, belonging to the Nepetoideae subfamily, is another genus with the same cosmopolitan distribution and

concentrating the largest distribution in Asia. Several descriptions of species of this genus are reported, however, they have quite different chemical substances from those found in the Scutellarioideae subfamily as we can verify the execution rate of the records of SOMs analyzed in clade III and clade IV [58].

3. Materials and Methods

3.1 Diterpenes Database

A database of diterpene molecules isolated from the Lamiaceae family was constructed based on a literature review that was performed using an electronic search in SciFinder (https://scifinder.cas.org/) and Web of Science (https://clarivate.com/products/web-of-science/), covering articles published between the years 1980 and 2017. Subsequently, the database was made available in the web tool SistematX [32]. The chemical structures, SMILES codes, names of the compounds (chemical and common), bibliographic references, as well as specific information for taxonomic classification (from family to species) and the geographical location of the species from which the compounds were isolated were compiled, and the total number was calculated instantaneously.

3.2 Molecular Descriptors

For all diterpene structures, SMILES codes were used as input data for Marvin and ChemAxon (http://www.chemaxon.com/). Then, Standardizer software (http://www.chemaxon.com/) was used to convert the various chemical structures into custom canonical representations, add hydrogens, aromatize, generate 2-D structures and save the compounds in SDF format. After processing in the Standardizer software, the 2-D structures of the compounds were used as input data in the Dragon 7.0 program [24]. This program has the capacity to calculate 5270 molecular descriptors covering several theoretical approaches and distributing the descriptors into 30 logical blocks. In Dragon 7.0, the coordinates of the atoms of each molecule were selected and then 301 molecular descriptors distributed in three blocks were calculated: ring descriptors, functional groups and atom centralizers [24]. Ring descriptors are numerical quantities that encode information about the presence of rings in a molecule;

functional groups are groups of atoms with characteristic and specific reactivity; centered descriptors are defined as the number of specific types of atoms in a molecule [59].

The constant variables were excluded for each block of descriptors and those that presented a different value in the series. The remaining 119 molecular descriptors that were submitted to statistical analysis were 32 rings, 39 functional groups and 37 atom-centered fragments.

In Dragon 7.0, the coordinates of the atoms of each molecule were selected and then 1024 fingerprints descriptors were calculated with the following atom options: atom type, aromaticity, attached hydrogens, connectivity (total), total bond order, connectivity (no H), ring memberships in smallest set of smallest rings (SSSR), smallest ring size in SSSR and bond order.

3.3 SVM and kNN Models

The Knime 3.6.2 software (Knime 3.4.0 the Konstanz Information Miner Copyright, 2003–2017, www.knime.org) was used to perform all of the following analyses. The descriptors and class variables were imported from the software Dragon 7.0, and for each one the data were divided using the “partitioning” node with the “stratified sample” option to create a training set and a test set, encompassing 80% and 20% of the compounds, respectively. Although the compounds were selected randomly, the same proportion of active and inactive samples was maintained in both sets. Two models were generated using the support vector machine algorithm (SVM) and the k-nearest neighbors algorithm (k-NN). The models were modeled following a 5-fold external cross-validation.

SVM is a supervised machine learning algorithm that analyzes data and recognizes patterns [60,61]. The parameters selected for the SVM for all generated models were polynomial, power 1.0, bias 1.0 and gamma 1.0.

k-NN consists of instance-based machine learning, i.e., the function is approximated only locally (neighbors) and the entire calculation is postponed until classification [62,63]. It is a technique that gives weight to neighbors’ contributions, so that the

nearest neighbors contribute more to the average than do the more distant ones [62,63]. The parameter selected for the SVM for all generated models was k = 3.

3.4 Self-Organizing Maps

The previously selected molecular descriptors were analyzed with SOM Toolbox 2.0 [37,64]. The SOM Toolbox is a set of MATLAB functions that can be used for the elaboration and implementation of neural networks because it contains functions for the creation, visualization and analysis of SOMs [37,64]. The data set was presented to the network before any adjustment was made. Subsequently, the data group was partitioned according to the regions of the map weight vectors at each training stage. Then, the correct prediction of these sets and the correct total prediction of the compounds were evaluated. In the most relevant models, the set was divided into training and testing to assess predictability. The training and test performances were evaluated by calculating the proportion of the number of samples classified correctly by the SOM. For each map, 5-fold cross-validation was performed, with data being partitioned into 80% training and 20% test (Tables 8 and 9). In the SOM, the sites containing molecules for each descriptor were identified to show existing chemical patterns. For the AD, which is defined as a theoretical region of the physicochemical and response space of the model that allows one to estimate the uncertainty in the prediction of a particular compound based on how similar it is to the training compounds employed in the model [60], the AD Enalos node in the Knime 3.7.1 software was used [61]. The AD based on the Euclidean distances was used to identify compounds in the test set for which predictions may be unreliable if the values are higher than AD = d+ Zσ, where d and σ are average Euclidian distance and standard deviation, respectively, of the set of samples in the training set that have lower Euclidian distance than the average values of all samples in the training set. The parameter Z is an empirical cut off value, 0.5 was used as the default.

Table 8. Summary of the five different training and test sets related to SOM obtained with diterpenes from clades III (Nep) and IV (Aju, Lam and Scu).

	Train Set		Test Set		Total
	Train	% Total	Test	% Total	
Clade III	2915	80	728	20	3643
Clade IV	1905	79.9	477	20.1	2382

Table 9. Summary of the five different sets of training and test related to SOM obtained with diterpenes only from Clade IV (Aju, Lam and Scu).

	Train Set		Test Set		Total
	Train	% Total	Test	% Total	
Aju	684	79.9	172	20.1	856
Lam	947	80	236	20	1183
Scu	273	79.8	69	20.2	342

4. CONCLUSIONS

The database of the present work presents a great diversity of diterpenes of the family Lamiaceae that were available in the web tool SistematX (https://sistematx.ufpb.br), with more than 4115 molecules distributed in 639 species of 66 genera and eight subfamilies, totaling more than 6386 botanical occurrences. The SOMs obtained from the Lamiaceae subfamilies, using molecular descriptors, separated the subfamilies with high accuracy rates (>80%) and corroborate previous phylogenetic studies by Li et al. [3]. Thus, SOMs based on physicochemical properties encoded from diterpenes are a useful tool to search for structures with defined characteristics and can be used, for example, in the search for diterpenes with potential biological activity using taxonomic and geographic data.

Author Contributions: Conceptualization, A.B.S.C., M.T.S., and L.S.; methodology, A.B.S.C., R.P.C.B., M.T.S., L.S., V.C.d.O.C., M.S.d.S. and J.F.T.; software, A.B.S.C., R.P.C.B., M.T.S.; validation, A.B.S.C., M.T.S., and L.S.; formal analysis, A.B.S.C., R.P.C.B., M.T.S.; investigation, A.B.S.C., M.T.S., and L.S.; data curation, A.B.S.C., R.P.C.B., M.T.S. and L.S.; writing—original draft preparation, A.B.S.C., R.P.C.B and M.T.S.

Funding: This research was funded by Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq), grant numbers 431254/2018-4 and 310919/2016-9. N.

Acknowledgments: We would like to thank the CNPq and Capes for financial Support.

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

REFERENCES

1. Casanova, L.M.; Costa, S.S. Synergistic Interactions in Natural Products: Therapeutic

Potential and Challenges. *Rev. Virtual Quím.* **2017**, *9*, 575–595. [CrossRef]

2. Singh, R. Chemotaxonomy: A Tool for Plant Classification. *J. Med. Plants Stud.* **2016**, *4*, 90–93.

3. Li, B.; Cantino, P.D.; Olmstead, R.G.; Bramley, G.L.C.; Xiang, C.-L.; Ma, Z.-H.; Tan, Y.-H.; Zhang, D.-X. A Large-Scale Chloroplast Phylogeny of the Lamiaceae Sheds New Light on Its Subfamilial Classification. *Sci. Rep.* **2016**, *6*, 34343. [CrossRef] [PubMed]

4. Basílio, I.J.L.D.; Agra, M.d.F.; Rocha, E.A.; Leal, C.K.A.; Abrantes, H.F. Comparative Pharmacobotanical Study of the Leaves of *Hyptis pectinata* (L.) Poit. and *Hyptis suaveolens* (L.) Poit. (Lamiaceae) [Estudo Farmacobotânico Comparativo Das Folhas de *Hyptis pectinata* (L.) Poit. e *Hyptis suaveolens* (L.) Poit. (Lamiaceae)]. *Acta Farm. Bonaer.* **2006**, *25*, 518–525.

5. Monteiro, F.K.d.S.; Pastore, J.F.B.; Melo, J.I.M.d. The Flora of Paraíba State, Brazil: Subfamilies Ajugoideae and *Viticoideae* (Lamiaceae). *Biota Neotrop.* **2018**, *18*. [CrossRef]

6. Frezza, C.; Venditti, A.; Serafini, M.; Bianco, A. Phytochemistry, Chemotaxonomy, Ethnopharmacology, and Nutraceutics of Lamiaceae. In *Studies in Natural Products Chemistry*; Elsevier: Cambridge, UK, 2019; Volume 62, pp. 125–178. [CrossRef]

7. Tundis, R.; Peruzzi, L.; Menichini, F. Phytochemical and Biological Studies of Stachys Species in Relation to Chemotaxonomy: A Review. *Phytochemistry* **2014**, *102*, 7–39. [CrossRef]

8. Sghaier, M.B.; Harizi, H.; Louhichi, T.; Krifa, M.; Ghedira, K.; Chekir-Ghedira, L. Anti-Inflammatory and Antiulcerogenic Activities of Leaf Extracts and Sesquiterpene from *Teucrium ramosissimum* (Lamiaceae). *Immunopharmacol. Immunotoxicol.* **2011**, *33*, 656–662. [CrossRef]

9. Piozzi, F.; Bruno, M.; Rosselli, S.; Maggio, A. The Diterpenoids from the *Genus hyptis* (Lamiaceae). *Heterocycles* **2009**, *78*, 1413. [CrossRef]

10. Yao, J.-L.; Fang, S.-M.; Liu, R.; Oppong, M.; Liu, E.-W.; Fan, G.-W.; Zhang, H. A Review on the Terpenes from Genus Vitex. *Molecules* **2016**, *21*, 1179. [CrossRef]

11. Martínez-Fructuoso, L.; Pereda-Miranda, R.; Rosas-Ramírez, D.; Fragoso-Serrano, M.; Cerda-García-Rojas, C.M.; Da Silva, A.S.; Leitão, G.G.; Leitão, S.G. Structure Elucidation, Conformation, and Configuration of Cytotoxic 6-Heptyl-5,6-Dihydro-2 H -Pyran-2-Ones from Hyptis Species and Their Molecular Docking to α-Tubulin. *J. Nat. Prod.* **2019**, *82*, 520–531. [CrossRef]

12. Dianita, R.; Jantan, I. Ethnomedicinal Uses, Phytochemistry and Pharmacological Aspects of the Genus Premna: A Review. *Pharm. Biol.* **2017**, *55*, 1715–1739. [CrossRef] [PubMed]

13. Benedec, D.; Hanganu, D.; Oniga, I.; Tiperciuc, B.; Olah, N.-K.; Raita, O.; Bischin, C.; Silaghi-Dumitrescu, R.; Vlase, L. Assessment of Rosmarinic Acid Content in Six Lamiaceae Species Extracts and Their Antioxidant and Antimicrobial Potential. *Pak. J. Pharm. Sci* **2015**, *28*, 2297–2303.

14. Shang, X.; He, X.; He, X.; Li, M.; Zhang, R.; Fan, P.; Zhang, Q.; Jia, Z. The Genus

Scutellaria an Ethnopharmacological and Phytochemical Review. *J. Ethnopharmacol.* **2010**, *128*, 279–313. [CrossRef] [PubMed]

15. Alvarenga, S.V.; Gastmans, J.P.; Rodrigues, G.d.V.; Moreno, P.R.; Emerenciano, V.d.P. A Computer-Assisted Approach for Chemotaxonomic Studies—Diterpenes in Lamiaceae. *Phytochemistry* **2001**, *56*, 583–595. [CrossRef]

16. Johnson, S.R.; Bhat, W.W.; Bibik, J.; Turmo, A.; Hamberger, B.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the *Mint family* (Lamiaceae). *J. Biol. Chem.* **2019**, *294*, 1349–1362. [CrossRef] [PubMed]

17. Hanson, J.R. Diterpenoids. *Nat. Prod. Rep.* **2000**, *17*, 165–174. [CrossRef] [PubMed]

18. Hanson, J.R. Diterpenoids of Terrestrial Origin. *Nat. Prod. Rep.* **2017**, *34*, 1233–1243. [CrossRef]

19. Hanson, J.R. Diterpenoids. *Nat. Prod. Rep.* **2009**, *26*, 1156–1171. [CrossRef]

20. Hanson, J.R. Diterpenoids. *Nat. Prod. Rep.* **1999**, *16*, 209–219. [CrossRef]

21. Kadereit, J.W. *The Families and genera of Vascular Plants*; Springer: Berlin/Heidelberg, Germany, 2004.

22. LI, B.; Olmstead, R.G. Two New Subfamilies in Lamiaceae. *Phytotaxa* **2017**, *313*, 222. [CrossRef]

23. Alves, V.; Braga, R.; Muratov, E.; Andrade, C. Quimioinformática: Uma Introdução. *Quim. Nova* **2017**, *41*, 202–212. [CrossRef]

24. Dragon 7.0. Kode Chemoinformatics. Available online: http://doi.wiley.com/10.1002/9783527628766 (accessed on 5 June 2019).

25. Scotti, L.; Tavares, J.F.; Silva, M.S.d.; Pessoa, J.; Emanuela, B.; Falcão, V.; Morais, L.; Cristina, G.; Soares, S.; Scotti, M.T. Chemotaxonomy of Three Genera of the Annonaceae Family Using Self-Organizing Maps and 13C NMR Data of Diterpenes. *Quim. Nova* **2012**, *35*, 2146–2152. [CrossRef]

26. Scotti, M.T.; Emerenciano, V.; Ferreira, M.J.; Scotti, L.; Stefani, R.; Da Silva, M.S.; Mendonça Junior, F.J. Self-Organizing Maps of Molecular Descriptors for Sesquiterpene Lactones and Their Application to the Chemotaxonomy of the Asteraceae Family. *Molecules* **2012**, *17*, 4684–4702. [CrossRef] [PubMed]

27. Zhang, Q.-Y.; Aires-de-Sousa, J. Structure-Based Classification of Chemical Reactions without Assignment of Reaction Centers. *J. Chem. Inf. Model.* **2005**, *45*, 1775–1783. [CrossRef]

28. Dacosta, F.; Terfloth, L.; Gasteiger, J. Sesquiterpene Lactone-Based Classification of Three Asteraceae Tribes: A Study Based on Self-Organizing Neural Networks Applied to Chemosystematics. *Phytochemistry* **2005**, *66*, 345–353. [CrossRef]

29. Hristozov, D.; Da Costa, F.B.; Gasteiger, J. Sesquiterpene Lactones-Based Classification of the Family Asteraceae Using Neural Networks and k -Nearest Neighbors. *J. Chem. Inf.*

Model. **2007**, *47*, 9–19. [CrossRef]

30. Wagner, S.; Hofmann, A.; Siedle, B.; Terfloth, L.; Merfort, I.; Gasteiger, J. Development of a Structural Model for NF-KB Inhibition of Sesquiterpene Lactones Using Self-Organizing Neural Networks. *J. Med. Chem.* **2006**, *49*, 2241–2252. [CrossRef]

31. Fernandes, M.B.; Scotti, M.T.; Ferreira, M.J.P.; Emerenciano, V.P. Use of Self-Organizing Maps and Molecular Descriptors to Predict the Cytotoxic Activity of Sesquiterpene Lactones. *Eur. J. Med. Chem.* **2008**, *43*, 2197–2205. [CrossRef]

32. Scotti, M.T.; Herrera-Acevedo, C.; Oliveira, T.; Costa, R.; Santos, S.; Rodrigues, R.; Scotti, L.; Da-Costa, F. SistematX, an Online Web-Based Cheminformatics Tool for Data Management of Secondary Metabolites. *Molecules* **2018**, *23*, 103. [CrossRef]

33. Valli, M.; Dos Santos, R.N.; Figueira, L.D.; Nakajima, C.H.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Development of a Natural Products Database from the Biodiversity of Brazil. *J. Nat. Prod.* **2013**, *76*, 439–444. [CrossRef]

34. DNP Database. Dictionary of Natural Products. Available online: http://dnp.chemnetbase.com/faces/ chemical/ChemicalSearch.xhtml (accessed on 5 June 2019).

35. Graham, J.G.; Farnsworth, N.R. The NAPRALERT Database as an Aid for Discovery of Novel Bioactive Compounds. In *Comprehensive Natural Products II*; Elsevier: Chicago, IL, USA, 2010; pp. 81–94. [CrossRef]

36. Marinlit Database. A database of the Marine Natural Products Literaturea Database Dedicated to Marine Natural Products. Available online: http://pubs.rsc.org/marinlit/introduction (accessed on 5 June 2019).

37. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. SOM Toolbox for Matlab 5. Available online: http://www.cis.hut.fi/projects/somtoolbox/ (accessed on 5 June 2019).

38. Zhan, R.; Li, X.N.; Du, X.; Wang, W.G.; Dong, K.; Su, J.; Li, Y.; Pu, J.X.; Sun, H.D. Bioactive Ent-Kaurane Diterpenoids from Isodon Rosthornii. *J. Nat. Prod.* **2013**, *76*, 1267–1277. [CrossRef] [PubMed]

39. Wu, H.Y.; Wang, W.G.; Jiang, H.Y.; Du, X.; Li, X.N.; Pu, J.X.; Sun, H.D. Cytotoxic and Anti-Inflammatory Ent-Kaurane Diterpenoids from Isodon Wikstroemioides. *Fitoterapia* **2014**, *98*, 192–198. [CrossRef] [PubMed]

40. Castro, A.; Coll, J.; Arfan, M. Neo-Clerodane Diterpenoids from Ajuga Bracteosa. *J. Nat. Prod.* **2011**, *74*, 1036–1041. [CrossRef] [PubMed]

41. Narukawa, Y.; Shimizu, N.; Shimotohno, K.; Takeda, T. Two New Diterpenoids from Plectranthus Nummularius BRIQ. *Chem. Pharm. Bull.* **2001**, *49*, 1182–1184. [CrossRef]

42. Chen, H.; Tan, R.-X.; Liu, Z.-L.; Zhao, C.-Y.; Sun, J. A Clerodane Diterpene with Antibacterial Activity from Ajuga Lupulina. *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **1997**, *53*, 814–816. [CrossRef]

43. Coll, J.; Tandrón, Y.A. Neo-Clerodane Diterpenoids from Ajuga: Structural Elucidation and

Biological Activity. *Phytochem. Rev.* **2007**, *7*, 25–49. [CrossRef]

44. Grace, M.H.; Cheng, D.M.; Raskin, I.; Lila, M.A. Neo-Clerodane Diterpenes from Ajuga Turkestanica. *Phytochem. Lett.* **2008**, *1*, 81–84. [CrossRef]

45. Hussain, H.; Ahmad, V.U.; Anwar, S.; Miana, G.A.; Krohn, K. Chemical Constituents of Scutellaria Linearis. *Biochem. Syst. Ecol.* **2008**, *36*, 490–492. [CrossRef]

46. Bruno, M.; Piozzi, F.; Rosselli, S. Natural and Hemisynthetic Neoclerodane Diterpenoids from Scutellaria and Their Antifeedant Activity. *Nat. Prod. Rep.* **2002**, *19*, 357–378. [CrossRef]

47. Maria, C.; Rodriguez, B.; Bruno, M.; Piozzi, F.; Savona, G.; Vassallo, N.; Servettaz, O. Neo-Clerodane diterpenoids from Scutellaria alpina. *Phytochemistry* **1995**, *38*, 181–187.

48. Ibrahim, S.A.; Ali, M.S. Constituents of Nepeta Crassifolia (Lamiaceae). *Turk. J. Chem.* **2007**, *31*, 463–470.

49. Yeon, E.T.; Lee, J.W.; Lee, C.; Jin, Q.; Jang, H.; Lee, D.; Ahn, J.S.; Hong, J.T.; Kim, Y.; Lee, M.K.; et al. Neo-Clerodane Diterpenoids from Scutellaria Barbata and Their Inhibitory Effects on LPS-Induced Nitric Oxide Production. *J. Nat. Prod.* **2015**, *78*, 2292–2296. [CrossRef] [PubMed]

50. Hennebelle, T.; Sahpaz, S.; Skaltsounis, A.L.; Bailleul, F. Phenolic Compounds and Diterpenoids from Marrubium Peregrinum. *Biochem. Syst. Ecol.* **2007**, *35*, 624–626. [CrossRef]

51. Takeda, Y.; Yanagihara, K.; Masuda, T.; Otsuka, H.; Honda, G.; Takaishi, Y.; Sezik, E.; Yesilada, E. Labdane Diterpenoids from Marrubium Globosum Ssp. Globosum. *Chem. Pharm. Bull.* **2011**, *48*, 1234–1235. [CrossRef] [PubMed]

52. Guo, P.; Li, Y.; Xu, J.; Liu, C.; Ma, Y.; Guo, Y. Bioactive Neo-Clerodane Diterpenoids from the Whole Plants of Ajuga Ciliata Bunge. *J. Nat. Prod.* **2011**, *74*, 1575–1583. [CrossRef] [PubMed]

53. Rodríguez, B.; De la Torre, M.C.; Jimeno, M.L.; Bruno, M.; Fazio, C.; Piozzi, F.; Savona, G.; Perales, A. Rearranged Neo-Clerodane Diterpenoids from Teucrium Brevifolium and Their Biogenetic Pathway. *Tetrahedron* **1995**, *51*, 837–848. [CrossRef]

54. Bao, H.; Zhang, Q.; Ye, Y.; Lin, L. Naturally occurring furanoditerpenoids: distribution, chemistry and their pharmacological activities. *Phytochem. Rev.* **2016**, *16*, 235–270. [CrossRef]

55. Fraga, B.M.; Bressa, C.; Fernandez, C.; Gonzalez, P.; Guillermo, R.; Hernández, M.G. Diterpenes from Sideritis Infernalis and S. Candicans. *Zeitschrift fur Naturforschung B* **2002**, *26*, 189–194. [CrossRef]

56. Fraga, B.M.; Hernández, M.G.; Fernández, C.; Santana, J.M.H. A Chemotaxonomic Study of Nine Canarian Sideritis Species. *Phytochemistry* **2009**, *70*, 1038–1048. [CrossRef]

57. Hanson, J.R. Diterpenoids. *Nat. Prod. Rep.* **1989**, *6*, 347–358. [CrossRef]

58. Sun, H.D.; Huang, S.X.; Han, Q.B. Diterpenoids from Isodon Species and Their Biological Activities. *Nat. Prod. Rep.* **2006**, *23*, 673–698. [CrossRef] [PubMed]

59. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009. [CrossRef]

60. Baskin, I.I. Machine Learning Methods in Computational Toxicology. In *Methods in Molecular Biology*; Humana Press Inc.: New York, NY, USA, 2018; Volume 1800, pp. 119–139. [CrossRef]

61. Mei, H.; Zhou, Y.; Liang, G.; Li, Z. Support Vector Machine Applied in QSAR Modelling. *Chin. Sci. Bull.* **2005**, *50*, 2291–2296. [CrossRef]

62. Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Amer. Statist.* **1992**, *46*, 175–185.

63. Cheng, D.; Zhang, S.; Deng, Z.; Zhu, Y.; Zong, M. kNN Algorithm with Data-Driven k Value. In *Lecture Notes in Computer Science*; Springer: Cambridge, UK, 2014; Volume 8933, pp. 499–512. [CrossRef]

64. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. Self−Organizing Map in Matlab: The SOM Toolbox. In Proceedings of the Matlab DSP Conference, Espoo, Finland, 16–17 November 1999; pp. 35–40.

Sample Availability: Samples of the compounds are not available from the authors.

Capítulo 3

Neste capítulo, reporta-se os constituintes químicos que foram isolados das partes aéreas de Leptohyptis macrostachys, os diterpenos: Hyptenol (1), eritroxilol B (2) e 3β-acetoxi-kaur-15-en-17-óico (6), e também os triterpenos: acetato de ácido oleanólico (3), ácido betulínico (4), ácido tormentico (5). Devido ao grande potencial antimicrobiano já reportado em espécies da família Lamiaceae, decidiu-se realizar avaliações *in silico* foram com os diterpenos 1 a 6. Para tanto foi construído um modelo preditivo, no qual diferentes tipos de descritores moleculares foram calculados para o banco de moléculas que apresentam atividade conhecida contra *Mycobacterium tuberculosis*. Os diterpenos 1, 2 e 6 foram então usados para investigar qualquer atividade antituberculose prevista. Os modelos foram gerados pelo algoritmo RF, e os descritores moleculares foram calculados usando o software Dragon 7. O modelo foi criado usando um método de validação externa.

Ao analisar o perfil de desempenho do modelo, excelentes taxas de previsão de atividade foram observadas com precisão de acordo com avaliações moleculares e, posteriormente, foi realizado as análises *in vitro* para determinar a suscetibilidade à micobactéria tuberculose. Sendo assim, os resultados demonstraram que o ácido *ent*-3β-acetoxi-kaur-15-en-17-óico (6) apresenta atividade biológica com MIC 62.5 µM contra *Mycobacterium tuberculosis* H37Ra.

A new labdane diterpene from the aerial segments of *Leptohyptis macrostachys* (L'H´erit.) Harley & J.F.B. Pastore

Andreza B.S. Cavalcanti[a], Pedro T.R. de Figueiredo [a], Carlos A.G. Veloso [a], Gabriela C. S. Rodrigues [a], Mayara dos S. Maia [a], Alex France Messias Monteiro [a], Valnês S. Rodrigues Junior [b], Ana P.O.T. Castelo-Branco [b], Maria de F. Agra [b], Raimundo B. Filho [c], Marcelo S. da Silva [a], Josean F. Tavares [a], Vicente C. de O. Costa [a], Luciana Scotti [a], Marcus T. Scotti [a,*]

[a] Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, Campus I, Cidade Universita´ria, João Pessoa, PB, Brazil
b Program of Biotechnology, Center for Biotechnology, Federal University of Paraíba, Campus I, Cidade Universitária, João Pessoa, PB, Brazil
c Laboratory of Chemical Science, State University of North Fluminense, Campos dos Goytacazes, Rio de Janeiro, RJ, Brazil

* Corresponding author.
E-mail address: mtscotti@gmail.com (M.T. Scotti).

Received 20 July 2020; Received in revised form 22 March 2021; Accepted 26 March 2021
Available online 6 April 2021

Artigo publicado na revista: **Phytochemistry Letters**, 2019, 43, 117-122.

https://doi.org/10.1016/j.phytol.2021.03.022

Fator de impacto: 1.679

ABSTRACT: A new labdane diterpene acid labda-8(17),14-dien-913-epoxy-12β-hydroxy-19-oic, denominated hyptenol (1), along with five other compounds: erythroxylol B (2), oleanolic acid acetate (3), betulinic acid (4), tormentic acid (5), and *ent*-3β-acetoxy-kaur-15-en-17-oic acid (6) were identified as chemical constituents of *Leptohyptis macrostachys* (aerial segments). The structures were identified through spectral data analysis using uni- and bi- dimensional nuclear magnetic resonance (NMR). Due to their great antimicrobial potential, in silico evalua- tions were performed with these compounds isolated from the Lamiaceae family. We constructed a predictive model, with molecular evaluations and then performed in vitro analyses to determine tuberculosis mycobacteria susceptibility. The results demonstrated that *ent*-3β-acetoxy-kaur-15-en-17-oic acid (6) presents significant biological activity against *Mycobacterium tuberculosis*.

Keywords: Lamiaceae; *Leptohyptis macrostachys*; Diterpenes; virtual screening; *Mycobacterium tuberculosis*

1. Introduction

More than 70 % of drugs recently approved for treatment against bacterial infections have come from natural products. For medicines, the Lamiaceae family is among the ten-most researched plant families (Sala-Carvalho et al., 2019). Lamiaceae (being herbaceous plants and shrubs) is composed of 258 genera and consists of approximately 7193 species (Aghakhani et al., 2018; Taamalli et al., 2015). In Brazil, 46 genera and 524 species are found (Soares et al., 2019), creating a great diversity of secondary metabolites, that include terpenoids (Fraga et al., 2009; Yao et al., 2016), such as mono, sesqui, di, and tri-terpenoids, along with steroids (Fraga et al., 2009), flavonoids (Dianita and Jan- tan, 2017; Fraga et al., 2009; He and Mu, 2015), coumarins (Olennikov et al., 2013), and alkaloids as well (Shang et al., 2014). Diterpenes are frequently extracted, including abietanes, labdanes, neo-clerodanes, primaranes, and *ent*-kaurenes (Jassbi et al., 2016; Piozzi et al., 2009; Zhang et al., 2018), and responsible for various biological activities.

Their antibacterial (Radulovic et al., 2010), antifungal (Cole et al., 1991; Msonthq, 2000; Zheng et al., 2012), antioxidant (Ba Vinh et al., 2018), anti-inflammatory (Huang

et al., 2015), and antiviral (Tada et al., 1994) activities have been well noted. Several Lamiaceae species have been evaluated for their biological potential including *Leptohyptis macrostachys* (L'H´erit.), Harley and J.F.B. Pastore (previously *Hyptis macrostachys* Benth.) which can be found in the semiarid northeast (is popularly known as "alfavaca-brava" or "hortelã-do-mato"), and is used in folk medicine to treat respiratory diseases (Agra et al., 2008; Harley and Pastore, 2012). Pharmacological studies have shown that extracts from this species display spasmolytic activity (de O.Costa et al., 2014; Sedano-Partida et al., 2020). Because of the great antimicrobial potential presented by Lamiaceae diterpenes, we evaluated aerial segments of the species *L. macrostachys* to identify bioactive compounds. Associating our study with natural products, we also sought to identify possible aminotransferase and *M. tuberculosis* InhA inhibitors. Aminotransferase is essential for survival of the bacteria and its inhibition causes intracellular metabolic amino acid disequilibrium (Castell et al., 2010). InhA contributes to the production of mycolic acids, which are important constituents of the mycobacterial cell wall (Marrakchi et al., 2000). Virtual Random Forest based screening using quantitative structure-activity relationships (QSAR), is an excellent strategy for predicting the anti-tuberculosis potential of selective inhibitors. We thus performed virtual screening, to identify diterpenes with promising activity against *Mycobacterium tuberculosis*, and then verified there in vitro activity as well.

2. Results and discussion

The phytochemical study was performed on the dichloromethane phase of a crude ethanolic extract of *L. macrostachys*, which produced six compounds (Fig. 1). Compound **1** was obtained as a white amorphous powder, with a melting point (m.p.) of 67–70 °C, $[\alpha]^{D}_{20}$ + 54 (c = 0.001, chloroform). IR (KBr): 3419− 2875 (− OH), 1693 (CO=). HR-EI-MS: 333.1885 ($[M–H]^-$, $C_{20}H_{30}O_4$). In the ^{13}C nuclear magnetic resonance (NMR) spectrum, 20 signals were observed, corresponding to 20 carbon atoms: 6 un-hydrogenated, 8 methylene, 3 methine, and 3 methyl. The signals at δ_C 47.9, 88.3, and 83.9 were attributed to C-5, C-9, and C-13, respectively, for the grindelane diterpenoids (Román et al., 2000). The signals at δ_C 153.5 and 107.3, which are characteristic of double exocyclic carbons, were respectively attributed to carbons

C-8 and C-17. The signal at δ_C 183.4 was consistent with a carboxylic acid at C-4, and the signal at δ_C 28.7 was attributed to the methyl equatorial C-19 (Meragelman et al., 2004). The signals at C-14, C-15, and C-16, in the grindelane diterpenoids, were observed at δ_C 145.2, 110.7, and 27.9, respectively (Román et al., 2000). The signal at δ_C 19.7 was attributed to methyl C-16 in the grindelane diterpenoids, with a double bond between C-14 and C-15, whose chemical shifts were respectively δ_C 143.2 and 112.6. The chemical shifts for C-16 and C-9 were protected due to the γ effect of an inserted β hydroxyl group, at the C-12 position, for which the chemical shift was observed at δ_C 78.0. The other shifts are shown in Table 1. In the ^{1}H NMR, spectrum the signals at δ_H 4.96 (brs, 1 H) and 4.76 (brs, 1 H), which are typical protons associated with the exocyclic double of labdane diterpenes, were attributed to the two protons at H-17 (Meragelman et al., 2004). The signals at δ_H 5.96 (^{1}H, dd, J = 11.0 and 17.5 Hz), δ_H 5.23 (^{1}H, dd, J = 1.0 and 17.5 Hz), and δH 5.06 (^{1}H, dd, J = 11.0 and 1.5 Hz) were attributed to H-14 and 2H-15. In addition, three singlets were observed at δ_H 0.66, 1.16, and 1.21, corresponding to three methyl groups, and δ_H 4.05 (t, J =7.0 Hz) was attributed to H-12. The other chemical shifts are shown in Table 1. The heteronuclear multiple quantum coherence (HMQC) spectrum revealed a correlation between δ_H 4.05 and δ_C 78.0, confirming the insertion of a hydroxyl group at C-12. In the heteronuclear multiple bond coherence (HMBC) spectrum, the following correlation signals were observed: δ_H 1.16 (CH_3-16) with the carbons at δ_C 78.0, 83.0, and 143.2, confirming respective attributions to C-12, C-13, and C-14; δ_H 0.66 (CH_3-20) with the carbons at δ_C 32.4, 47.5, 88.3, and 43.2, respectively corresponding to C-1, C-5, C-9, and C-10; δ_H 2.18 (H-11) with carbons at δ_C 153.5, 88.3, 78.0, and 83.9, confirming H-11, and the correlation between this signal and the carbon at δ_C 34.6 confirmed at C-11. These correlations confirmed the presence of a labdane diterpenic skeleton. The correlation between δ_H 1.21 (CH_3−18) and the carbons δ_C 37.8, 44.2, and 183.1, corresponding to C-3, C-4, and C-19, respectively, confirmed the presence of a carboxyl at C-19. The other correlations are compiled in Table 1. In the ^{1}H-^{1}H correlation spectroscopy (COSY) spectrum, signal correlations at δ_H 5.96 (H-14) were observed with δ_H 5.23 and 5.06 (2H-15), confirming a monosubstituted double bond. In the nuclear Overhauser effect spectroscopy (NOESY) spectrum, the following correlations were observed: δ_H 2.10 (H-7eq) with δ_H 4.96 (H-17); δ_H 1.02 (H-3ax) with

δ_H 1.94 (H-5); δ_H 194 (H-5) with δ_H 1.21 (CH_3-18); δ_H 0.66 (CH_3-20) with δ_H 2.18 (H-11) and δ_H 1.82 (H-2ax); and δ_H 122 (H-1eq) with δ_H 4.05 (H-12) (Fig. 2). The stereochemistry of methyl in CH_3- 20 was defined as beta from the positive record of $[\alpha]^{D}_{20}$. The chemical displacement of (CH_3-18) at 28 ppm infers that it is trans in relation to (CH_3-20), so this methyl was defined as alpha and beta carboxylic acid. The NOESY of (CH_3-18) with H-5 defines five alphas. The NOESY of (CH_3-20) with 2H-11 defines stereochemistry at 9 inferring alpha oxygen, and 2H-11 with (CH_3-16) confirms the beta stereochemistry of (CH_3-16). Chemical displacement of this methyl when compared to grindelic acid creates a protective effect. Thus, the stereochemistry of the hydroxyl in C-12 was defined as betas. Chemical calculations based on density functional theory (DFT) can be used to provide reliable predictions of NMR parameters (^{1}H and ^{13}C NMR chemical shifts). Initially, based on the NMR results, two possible structures could be defined for compound 1 to justify protection at CH_3-16. The combination of NMR data with DFT was therefore important for defining the structure, which was confirmed by comparisons between the NMR experimental data and the NMR results from the theoretical modeling. After analyzing these data, compound 1 a new natural product, denominated hyptenol was determined: labda-8(17),14-dien-9,13-epoxy-12β-hydroxy-19-oic acid.

The other compounds were identified using direct comparisons with their spectroscopic data, which had been previously described in the literature: erythroxylol B (2) (Abad et al., 1994), oleanolic acid acetate (3) (Brochini et al., 1994), betulinic acid (4) (Bisoli et al., 2008), tormentic acid (5) (Jang et al., 2006), and *ent*-3β-acetoxy-kaur-15-en-17-oic acid (6) (Dutra et al., 2014).

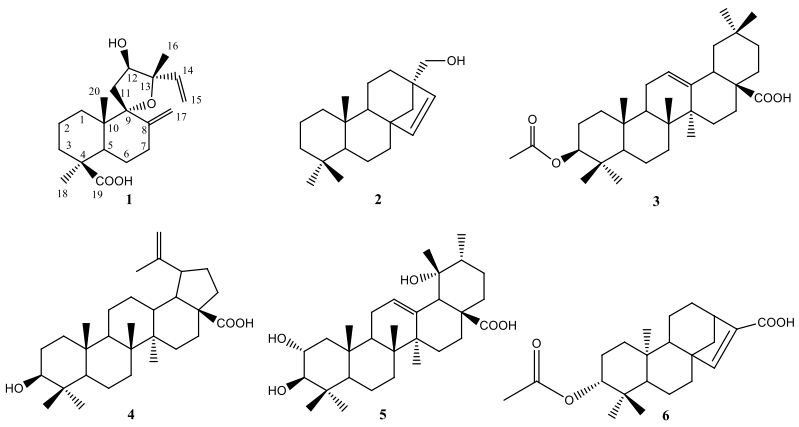


Figure 1. Chemical constituents isolated from *L. macrostachys*.

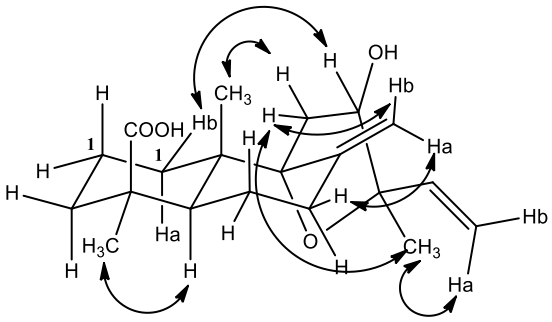


Figure 2. NOESY correlations of 1.

Table 1. NMR data for 1 in 500 MHz, $CDCl_3$.

Position	δ_C	δ_H	HMBC	COSY
1	32.4	1.68 (ddd, *J* = 4.0, 13.5, 4.0, Hz, H1 ax) 1.22 (m, H1 eq)		H-1 eq
2	19.6	1.82 (m, H2 ax), 1.51 (m, H2 eq)	C-1	H-3, H-1
3	37.8	1.02 (ddd, *J* = 13.5, 4.0, 4.0 Hz, H3 ax), 2.10 (m, H3 eq)		H-2, H-3
4	44.2	-		
5	47.9	1.94 (dd, *J* = 13.0, 3.5 Hz)		H-6
6	25.3	1.90 (m, H6 ax), 1.30, (m, H6 eq)	C-8	
7	33.3	2.26 (ddd, *J* = 5.5, 6.0, 13.0 Hz, H7 ax), 2.10 (m, H7 eq)	C-6, C-8	H-7, H-6
8	153.5	-		
9	88.3	-		
10	43.2	-		
11	34.6	2.18 (d, *J* = 5.0 Hz)	C-9, C-12, C-13	
12	78.0	4.05 (t, *J* = 7.0 Hz)	C-9, C-14, C-16	
13	83.9	-		
14	143.2	5.96 (dd, *J* = 11.0, 17.5 Hz)	C-13	H-15a, 15b
15	112.6	5.23 (dd, *J* = 1.0, 17.5 Hz), 5.06 (dd, *J* = 1.5, 11.0 Hz)	C-13, C-14	H-14
16	19.5	1.16 (s),	C-12, C-13,	
17	107.3	4.96 (sl), 4.76 (sl)	C-7, C-8, C-9	
18	28.7	1.21 (s)	C-4, C-5, C-19	
19	183.1	-		
20	15.3	0.66 (s)	C-1, C-5, C-9	H-1 eq

To construct the Random forest (RF) model, different types of molecular descriptors were calculated for the bank of molecules presenting known activity against *Mycobacterium tuberculosis*. Diterpenes 1, 2, and 6 were then used to investigate any predicted antituberculosis activity. The models were generated by the RF algorithm, and the molecular descriptors were calculated using Dragon 7 software. The model was created using an external validation method. Upon analyzing the model's performance profile, excellent activity forecasting rates were observed with accuracies

of greater than 89 %. The classification rate of the model was evaluated based on the receiver operating characteristic (ROC) curve, yielding 0.95 for both the test and cross-validation datasets. The Matthews correlation coefficient (MCC) was used to assess the classification of the model, which presented values of 0.80 for the test set and 0.78 for the validation set. The study models thus generated excellent RF method classification rates. After assessing the performances of the models, we investigated the biological anti-tuberculosis activity predictions for the diterpenes identified. Molecule 6 presented a probability of activity at above 52 % and was considered by the model to be active. Of the 15 descriptors that most contributed to classification of the compounds as active, seven were information index descriptors (IC1, IC3, TIC2, TIC3, IDE, IVDE, and HDcpx), calculated as the information content of the molecules, and based on equivalence class calculations from the molecular graphs, e.g. neighborhood symmetry indices. Two descriptors were also observed based on the matrix's 2D; referring to the eigenvector coefficient sum of the last eigenvector (VEA1) and the eigenvector coefficient sum of the polarizability weighted distance matrix (VEp1), a geometric descriptor (SPAM), two connectivity descriptors (X5A and XMOD), and a 2D auto correlation descriptor (GATS8p). According to the coupling studies, the following important residues for Inhibin alpha (InhA) were considered: Phe149, Tyr158, Pro193, Ile194, Arg195, Met199, Leu218, Glu219, and Met232. We noticed that all of the diterpenes could interact at the active site of the enzyme through the various hydrophobic bonds with the amino acids mentioned above. Hydrogen bonds were observed between compound 1 and amino acid Ile194; between compound 2 and amino acid Pro156; and between compound 6 and amino acids Ile194 and Glu219. Analyzing residues important for aminotransferase inhibition, all of the diterpenes were able to interact at the active site of the enzyme through electrostatic bonds. Hydrogen bonds were observed between compound 1 and amino acid Ser233; between compound 2 and amino acids Asp202 and ASN174; and between compound 6 and amino acids Arg242, Ser101, Asn174, and Pro172.

Both complex flexibility and conformational changes occurring during the molecular dynamics simulations were studied using diterpene 2 and 6 interactions (with their crystallographic ligands). The rootmean-square deviation (RMSD) analysis of aminotransferase revealed that the protein (within 10 ns) achieved conformations

ranging from 0.2 to 0.8 nm in size, demonstrating significant instability. In contrast with the non-complexed protein, the protein when complexed with the diterpene and the inhibitor ligand presented its greatest stability after 5 ns,. When comparing ligand flexibility, we found that compound 2 displayed very small peaks of instability, measured in angstroms. The instability of the PDB ligand was greater. We therefore suggest that although the protein undergoes minor structural changes and the compound has peaks of instability, the molecule tends to remain within the active site. This suggests that its binding affinity is not lost, even in the presence of potentially interfering factors, such as temperature, pressure, solvents, and ions. For the InhA enzyme, RMSD analysis showed that the protein reached conformations ranging from 0.2 to 0.45 nm in size, within 10 ns, presenting low instability. The protein when complexed with both the diterpene and the inhibitor ligand presented increased stability after 5 ns, unlike the non-complexed protein. When comparing the flexibility of the ligands, we found that compound 6 presented greater stability than the PDB ligand (greater peaks of instability were observed for the PDB ligand). We therefore suggest that the binding affinity of compound 6 is stable, with bonds and interactions that are difficult to break, even in the presence of differing factors, such as temperature, pressure, solvents, and ions.

Based on the results, we verified the potential antitubercular activity of compounds 1 and 6 by performing *in vitro* tests against *Mycobacterium tuberculosis* H37Ra. The results can be found in Table 2. We lacked a sufficient quantity of Sample 2 for adequate testing.

Table 2. Activity of compounds against *Mycobacterium tuberculosis* H37Ra.

Compounds	**MIC (μM)[a]**
	Mycobacterium tuberculosis
INH[b]	3.8
RIF[b]	0.02
MOX[b]	0.2
1	>500
6	62.5

As shown in Table 2, diterpene 1 did not present activity against *M. tuberculosis*, however diterpene 6 presented a minimum inhibitory concentration (MIC) of 62.5 μM. These biological test results, in which only diterpene 6 presented promising activity against *M. tuberculosis*, corroborated our in silico study and validated our

antitubercular activity prediction model. According to Soares and collaborators, certain kaurane and trachylobane diterpenes demonstrate anti-tuberculosis activity, with lower MIC values than the second-choice anti-TB drug - Cycloserine (122.4–498 μM) (Soares et al., 2017). We therefore, suggest that the active diterpene identified in this study is also more potent than Cycloserine.

3. Procedures

3.1 General experimental procedures

Melting points were determined with a digital apparatus model MQAPF-302, Microchemical (Araucaria, SC, Brazil). Infrared (IR) spectra were recorded on a BOMEM-MB 100 spectrophotometer (Quebec City, Quebec, Canada). ^{1}H and ^{13}C and two-dimensional NMR analyses were performed on a Varian-System spectrometer (Palo Alto, CA, USA), operating at 500 MHz and 125 MHz. $CDCl_3$ and pyridine-d_5 (Cambridge Isotope Laboratories, Tewksbury, MA, USA) were used as solvents, with TMS as an internal standard. High-resolution electrospray ionization mass spectrometry (HR-ESI-MS) was performed using a micrOTOF-II system, from Bruker (HRESIMS, Bruker, Billerica, MA, USA. Silica gel 60 (0.063− 0.20 and 0.04−0.063 mm; (ART 7734, Merck, Darmstadt, Germany) was used for column chromatography (CC), and silica gel thin-liquid chromatography (TLC) plates PF254 7749 (Merck) stained with iodine vapor and viewed under ultraviolet (UV) light (254/366 nm) were used to monitor the chromatographic purification procedures. The structures of the compounds were drawn using Marvin, 19.27.0 (2019, ChemAxon, http://www.chemaxon.com), Standardizer software was used to canonize the structures (Imre et al., 2003).

Geometric optimizations and conformational searches were performed using Spartan for Windows 10.0 (www.wavefun.com/products/windows/SpartanModel/win_model.html). The geometries of the compounds' chemical structures were initially optimized with Merck Molecular Force Field (MMFF) (Halgren, 1996), and a new geometric optimization was then performed, based on the semi-empirical method Austin Model 1 (AM1) (Dewar et al., 1985). A systematic search method was applied to analyze conformers and to select the conformers with the lowest minimum energy, using the AM1 and a Monte-Carlo algorithm (Metropolis and

Ulam, 1949). The lowest minimum energies were then selected and optimized, based on a DFT vibrational mode calculation. The DFT calculations were performed using Spartan 10 for Windows (Wavefunction, Irvine, CA, USA) (Lin et al., 2004). Each structure studied was examined at the EDF2/6− 311G* level, and the lowest energy structures were selected for calculations. The global minimum potential energy surface was used to determine each geometry, and ^{1}H and ^{13}C NMR chemical shifts were generated.

3.2 Plant material

Aerial segments of *L. macrostachys* were collected from Matureia, a municipality in Paraiba, Brazil), in August 2009, by Prof. Maria de Fátima Agra, PhD., at the Taxonomy & Pharmacobotany Laboratory of the Postgraduate Program in Natural and Synthetic Bioactive Products at the Federal University of Paraiba (UFPB). An exsiccata was deposited at the Prof. Lauro Pires Xavier Herbarium, at the UFPB Campus of João Pessoa, under the number 6947.

3.3 Extraction and isolation

The dried and pulverized plant material (2 kg) was extracted with EtOH (95 %). The ethanolic extract was then concentrated, yielding 200 g of crude ethanolic extract. A portion (100 g) was then subjected to vacuum filtration, with silica gel, hexane, dichloromethane, and ethyl acetate; and yielding hexane (1 g), dichloromethane (12.3 g), and ethyl acetate (18.7 g) phases. The dichloromethane phase was subjected to CC, using hexane:EtOAc (9:1, 7:3, 1:1, 3:7, 1:9), and resulting in 305 fractions, which were pooled after analytical TLC analysis to form ten fractions (A1–A10). The A2 fraction (4.0 g) was re-chromatographed on a column containing silica gel and eluted with hexane:EtOAc (9:1, 8:2, and 7:3). This resulted in 96 fractions, which were also pooled after analytical TLC analyses, to generate five fractions (B1–B5). The B3 fraction contained 22 mg of compound 1. The B2 fraction (2.0 g) was rechromatographed on a column containing silica gel and eluted with hexane:EtOAc (9:1, 8:2, and 7:3). This resulted in 56 fractions, which were pooled after analytical TLC analysis to generate seven fractions (C1–C7). The C3 fraction contained compound 2 (54 mg). The C4 fraction, eluted with hexane: EtOAc 2:8 (60 mg) was recrystallized

with methanol to obtain compound 3 (48 mg). The C5 fraction (70 mg) was recrystallized with methanol to obtained compound 4 (26 mg). The C6 fraction (100 mg) was re-chromatographed on a column containing silica gel and eluted with hexane:EtOAc (9:1, 8:2, and 7:3) and EtOAc, resulting in 56 fractions, which were pooled after analytical TLC analysis, to obtain four fractions (D1-D4). The D2 fraction contained compound 5 (15 mg). The D3 fraction (15 mg) was recrystallized with methanol to obtain compound 6 (12 mg). Then, all of the compounds and their species' geographic locations were made available at SistematX (https://sistematx.ufpb.br) (Scotti et al., 2018).

3.4 Cheminformatics

To construct a predictive model, a set of molecules with known activity against *Mycobacterium tuberculosis* (CHEMBL360) was obtained from the ChEMBL database. All of the compounds were carefully curated and standardized with duplicates and outliers removed for chemical and biological data in accordance procedures described by Fourches et al. (Fourches et al., 2016). Three-dimensional structures were generated using ChemaxonStandardiser v.18.17.0, (www.chemaxon.org). Chemical structures were classified as active or inactive. The bank contained 8487 molecules that were classified using MIC and –log MIC (pMIC) values. Compounds with pMIC ≥ 4.7 were considered active, yielding 4181 active molecules. Compounds with pMIC < 4.7 were considered inactive, yielding 4306 inactive molecules. The Knime 3.5.3 software (KNIME 3.5.3, Konstanz Information Miner Copyright, 2018, www. knime.org) was used to perform the analyses and to generate the in silico models. The Dragon 7.0 software (Mauri et al., 2006), was used to generate compound descriptors. The Random Forest (RF) algorithm, using WEKA nodes (Hall et al., 2021), was used to build predictive models. The banks of molecules and their calculated descriptors were imported from the Dragon software, and the data were divided into a "Partitioning" tool, using the "Stratified sample" option, which separated the data into Training and Test sets, which respectively represented 80 % and 20 % of all compounds. The sets were randomly selected, but the active/inactive substance proportions were maintained in both databases.

The external performances of the selected models were analyzed for sensitivity (true positive rate, which reflects the active rate), specificity (true negative rate, which reflects the inactivity rate), and for accuracy (general predictability). ROC curve sensitivity and specificity values have been suggested as more clearly describing actual performance than other metrics.

The models were validated using the Matthews coefficient, which allows the evaluation of a model based on the results obtained in the confusion matrix (Scotti et al., 2011). The applicability domain (APD) was used to validate the model, assessing whether its predictions were reliable (Scotti et al., 2011).

The diterpenes isolated in the study were also evaluated using virtual screening to identify molecules with potential activity against *M. tuberculosis*.

3.5 Molecular docking

Proteins together with their respective ligands were obtained from the Protein Data Bank (https://www.rcsb.org/) under the PDB IDs: 5YHV (aminotransferase), and 5VRN (InhA). The molecular docking study was performed using Molegro Virtual Docker, 6.0.1 software, (Molegro ApS, Aarhus, Denmark) (Thomsen and Christensen, 2006). Molegro was used to calculate the RMSD values for the PDB binding conformations, as generated by redocking in each docking program used in this study. For the ligand - enzyme coupling procedure, a GRID with a radius of 15 Å and a resolution of 0.30 was used to cover the binding site, defined using a known ligand for each enzyme. The model was designed to perform the adjustments based on the expected characteristics of the ligand and the enzyme, and using a heuristic search algorithm that combines the differential evolutions and uses the crystallographic ligand as a model. The cavity prediction algorithm (Moldock) and the Moldock score function were selected (Neves et al., 2018).

3.6 Molecular dynamics simulations

Molecular dynamics simulations were performed using GROMACS 5.0 software (Abraham et al., 2015; Berendsen et al., 1995). The ligand and protein topologies were prepared using the GROMOS96 54a7 force field. The molecular dynamics simulation was performed using the simple point charge (SPC) water model, based on the point

load extended in a cubic box (van Bondi, 1964). The system was neutralized by addition of ions (Cl^- and Na^+) in 5,000,000 steps, over 10 ns. The structure's flexibility was determined by RMSD analysis. The RMSD graphics were generated using Grace software (http://plasma-gate.we izmann.ac.il/Grace/).

3.7 Mycobacterial susceptibility assay

The control drugs rifampicin (RIF) and moxifloxacin (MFX) were purchased from Sigma-Aldrich, and isoniazid (INH) was obtained from ACROS Organics. *Mycobacterium tuberculosis* (H37Ra) was cultured, as previously described (Muradas ´ et al., 2018; Taneja and Tyagi, 2007). The MIC for each compound tested was determined in a 96-well, U-bottom, polystyrene microplate (Kumari et al., 2018). RIF, MFX, and INH solutions were prepared in dimethylsulfoxide (DMSO), at 1 mM concentrations. The isolated diterpenes were solubilized in DMSO, at 20 mM concentrations, and were then diluted in Middlebrook 7H9 broth, to achieve final concentrations of 500 μM. Serial dilutions were performed in 96-well U-bottom microplates, maintaining a final DMSO concentration of 2.5 %, for all wells. The mycobacterial suspensions were diluted in 7H9 medium, at a theoretical optical density (OD 600 nm) of 0.003, and incubated for 7 days at 37 ◦C (Giacobbo et al., 2017). Then, 30 μL of 0.02 % resazurin solution was added to analyze bacterial growth as based on resazurin reduction.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

The present work was carried out with the support the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq), grant numbers 309648/2019-0, 465536/2014-0, 431254/2018-4, and 404440/2016-9, and by the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.phytol.2021.03.022.

References

Abad, A., Agulló, C., Arnó, M., Marín, M.L., Zaragozá, R.J., 1994. Synthesis of C-17-functionalized beyerane diterpenes. Synthesis of (–)-erythroxylol B, (–)-erythroxydiol A and (–)-benuol. J. Chem. Soc. Perkin Trans. 1, 2987–2991.

Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., Lindah, E., 2015. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25. https://doi.org/10.1016/j. softx.2015.06.001.

Aghakhani, F., Kharazian, N., Lori Gooini, Z., 2018. Flavonoid constituents of Phlomis(lamiaceae) species using liquid chromatography mass spectrometry. Phytochem. Anal. 29, 180–195.

Agra, M.D.F., Silva, K.N., Basílio, I.J.L.D., de Freitas, P.F., Barbosa-Filho, J.M., 2008. Survey of medicinal plants used in the region Northeast of Brazil. Rev. Bras. Farmacogn. 18, 472–508. https://doi.org/10.1590/S0102-695X2008000300023.

Ba Vinh, L., Thi Minh Nguyet, N., Young Yang, S., Hoon Kim, J., Thi Vien, L., Thi Thanh Huong, P., Van Thanh, N., Xuan Cuong, N., Hoai Nam, N., Van Minh, C., 2018. A new rearranged abietane diterpene from Clerodendrum inerme with antioxidant and cytotoxic activities. Nat. Prod. Res. 32, 2001–2007.

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1995. GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56.

Bisoli, E., Garcez, W.S., Hamerski, L., Tieppo, C., Garcez, F.R., 2008. Bioactive pentacyclic triterpenes from the stems of Combretum laxum. Molecules 13, 2717–2728.

Brochini, Cudia B., Martins, D., Roque, N.F., da S.Bolzani, V., 1994. An oleanane acid from ALIBERTIA-EDULIS. Phytochemistry 1293–1295.

Castell, A., Mille, C., Unge, T., 2010. Structural analysis of mycobacterial branched-chain aminotransferase: implications for inhibitor design. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 549–557. https://doi.org/10.1107/S0907444910004877.

Cole, M.D., Bridge, P.D., Dellar, J.E., Fellows, L.E., Cornish, M.C., Anderson, J.C., 1991. Antifungal activity of neo-clerodane diterpenoids from Scutellaria. Phytochemistry 30, 1125–1127.

de O. Costa, V.C., Tavares, J.F., Silva, A.B., Duarte, M.C., Agra, M., de, F., Barbosa- Filho, J.M., de Souza, I.L.L., da Silva, B.A., Silva, M.S., 2014. Hyptenolide, a new α-pyrone with spasmolytic activity from Hyptis macrostachys. Phytochem. Lett. 8, 32–37.

Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P., 1985. Development and use of quantum mechanical molecular models. 76. AM1: a new general-purpose quantum mechanical molecular model. J. Am. Chem. Soc. 107, 3902–3909.

Dianita, R., Jantan, I., 2017. Ethnomedicinal uses, phytochemistry and pharmacological aspects of the genus Premna: a review. Pharm. Biol. 55, 1715–1739. https://doi.org/10.1080/13880209.2017.1323225.

Dutra, L.M., Bomfim, L.M., Rocha, S.L.A., Nepel, A., Soares, M.B.P., Barison, A., Costa, E. V., Bezerra, D.P., 2014. Ent-Kaurane diterpenes from the stem bark of Annona vepretorum (Annonaceae) and cytotoxic evaluation. Bioorg. Med. Chem. Lett. 24, 3315–3320.

Fourches, D., Muratov, E., Tropsha, A., 2016. Trust, but verify II: a practical guide to chemogenomics data curation. J. Chem. Inf. Model. 56, 1243–1252. https://doi.org/10.1021/acs.jcim.6b00129.

Fraga, B.M., Hern´andez, M.G., Fern´andez, C., Santana, J.M.H., 2009. A chemotaxonomic study of nine Canarian Sideritis species. Phytochemistry 70, 1038–1048. https://doi.org/10.1016/j.phytochem.2009.05.011.

Giacobbo, B.C., Pissinate, K., Rodrigues-Junior, V., Villela, A.D., Grams, E.S., Abbadi, B. L., Subtil, F.T., Sperotto, N., Trindade, R.V., Back, D.F., Campos, M.M., Basso, L.A., Machado, P., Santos, D.S., 2017. New insights into the SAR and drug combination synergy of 2-(quinolin-4-yloxy) acetamides against Mycobacterium tuberculosis. Eur. J. Med. Chem. 126, 491–501. https://doi.org/10.1016/j.ejmech.2016.11.048.

Halgren, T.A., 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comput. Chem. 17, 490–519.

Hall, M., National, H., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., n.d. The WEKA Data Mining Software: An Update 11, 10–18.

Harley, R.M., Pastore, J.F.B., 2012. A generic revision and new combinations in the Hyptidinae (Lamiaceae), based on molecular and morphological evidence. Phytotaxa 58, 1–55. https://doi.org/10.11646/phytotaxa.58.1.1.

He, J.-M., Mu, Q., 2015. The medicinal uses of the genus Mahonia in traditional Chinese medicine: an ethnopharmacological, phytochemical and pharmacological review. J. Ethnopharmacol. 175, 668–683. https://doi.org/10.1016/j.jep.2015.09.013.

Huang, Z., Zhu, Z.-X., Li, Y.-T., Pang, D.-R., Zheng, J., Zhang, Q., Zhao, Y.-F., Ferreira, D., Zjawiony, J.K., Tu, P.-F., Li, J., 2015. Anti-inflammatory labdane diterpenoids from Leonurus macranthus. J. Nat. Prod. 78, 2276–2285. https://doi.org/10.1021/acs.jnatprod.5b00635.

Imre, G., Veress, G., Volford, A., Farkas, O¨ ., 2003. Molecules from the Minkowski space: an approach to building 3D molecular structures. J. Mol. Struct. THEOCHEM 666, 51–

59. Jang, D.-S., Kim, J.-M., Lee, G.-Y., Kim, J.-H., Kim, J.-S., 2006. Ursane-type triterpenoids from the aerial parts of Potentilla discolor. J. Appl. Biol. Chem. 49, 48–50.

Jassbi, A.R., Zare, S., Firuzi, O., Xiao, J., 2016. Bioactive phytochemicals from shoots and roots of Salvia species. Phytochem. Rev. 15, 829–867. https://doi.org/10.1007/ s11101-015-9427-z.

Kumari, M., Tiwari, N., Chandra, S., Subbarao, N., 2018. Comparative analysis of machine learning based QSAR models and molecular docking studies to screen potential anti-tubercular inhibitors against InhA of mycobacterium tuberculosis. Int. J. Comput. Biol. Drug Des. 11, 209–235.

Lin, C.Y., George, M.W., Gill, P.M.W., 2004. EDF2: a density functional for predicting molecular vibrational frequencies. Aust. J. Chem. 57, 365–370.

Marrakchi, H., Lane´elle, G., Quemard, A., 2000. InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. Microbiology 146, 289–296. https://doi.org/10.1099/00221287-146-2-289.

Mauri, A., Consonni, V., Pavan, M., Todeschini, R., 2006. DRAGON software: an easy approach to molecular descriptor calculations. Match 56, 237–248.

Meragelman, T.L., Pedrosa, D.S., Gil, R.R., 2004. Diterpenes from Stevia gilliesii. Biochem. Syst. Ecol. 32, 45–53.

Metropolis, N., Ulam, S., 1949. The monte carlo method. J. Am. Stat. Assoc. 44, 335–341.

Msonthq, J.D., 2000. 164. uncinatone, a new antifungal hydroquinone diterpenoid from.Helv. Chim. Acta 68.

Muradás, T.C., Abbadi, B.L., Villela, A.D., Macchi, F.S., Bergo, P.F., De Freitas, T.F., Sperotto, N.D.M., Timmers, L.F.S.M., De Souza, O.N., Picada, J.N., Fachini, J., Da Silva, J.B., De Albuquerque, N.C.P., Habenschus, M.D., Carrao, D.B., Rocha, B.A., Junior, F.B., De Oliveira, A.R.M., Mascarello, A., Neuenfeldf, P., Nunes, R.J., Morbidoni, H.R., Campos, M.M., Basso, L.A., Rodrigues-Junior, V.S., 2018. Pre- clinical evaluation of quinoxaline-derived chalcones in tuberculosis. PLoS One 13, 1–13. https://doi.org/10.1371/journal.pone.0202568.

Neves, B.J., Braga, R.C., Melo-Filho, C.C., Moreira-Filho, J.T., Muratov, E.N., Andrade, C. H., 2018. QSAR-based virtual screening: advances and applications in drug discovery. Front. Pharmacol. 9, 1–7. https://doi.org/10.3389/fphar.2018.01275.

Olennikov, D.N., Chirikova, N.K., Okhlopkova, Z.M., Zulfugarov, I.S., 2013. Chemical composition and antioxidant activity of tánara Óto´ (*Dracocephalum palmatum* Stephan), a medicinal plant used by the north-yakutian nomads. Molecules 18, 14105–14121. https://doi.org/10.3390/molecules181114105.

Piozzi, F., Bruno, M., Rosselli, S., Maggio, A., 2009. The diterpenoids from the genus Hyptis (Lamiaceae). Heterocycles 78, 1413. https://doi.org/10.3987/REV-08-651.

Radulovíc, N., Denı´c, M., Stojanovi´c-Radi´c, Z., 2010. Antimicrobial phenolic abietane diterpene from Lycopus europaeus L. (Lamiaceae). Bioorg. Med. Chem. Lett. 20, 4988–4991.

Román, L.U., Cambrón, J.I., Del Río, R.E., Hernández, J.D., Cerda-García-Rojas, C.M., Joseph-Nathan, P., 2000. Grindelane diterpenoids from Stevia subpubescens. J. Nat. Prod. 63, 226–229.

Sala-Carvalho, W.R., dos Santos, K.P., Sedano-Partida, M.D., da Silva-Luz, C.L., Pena Ferreira, M.J., Furlan, C.M., 2019. Is Hyptis lacustris A. St.-Hil. Ex Benth. (Lamiaceae) extract a good candidate for antibacterial uses? Ind. Crops Prod. 133, 26–32. https://doi.org/10.1016/j.indcrop.2019.02.054.

Scotti, L., Fernandes, M.B., Muramatsu, E., Pasqualoto, K.F.M., de P. Emereciano, V., Tavares, L.C., da Silva, M.S., Scotti, M.T., 2011. Self-organizing maps and VolSurf approach to predict aldose reductase inhibition by flavonoid compounds. Brazil. J. Pharmacogn. 21, 170–180. https://doi.org/10.1590/S0102-695X2011005000028.

Scotti, M.T., Herrera-Acevedo, C., Oliveira, T.B., Costa, R.P.O., De Oliveira Santos, S.Y. K., Rodrigues, R.P., Scotti, L., Da-Costa, F.B., 2018. SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. Molecules 23,1–10. https://doi.org/10.3390/molecules23010103.

Sedano-Partida, M.D., dos Santos, K.P., Sala-Carvalho, W.R., Silva-Luz, C.L., Furlan, C. M., 2020. A review of the phytochemical profiling and biological activities of Hyptis Jacq.: a Brazilian native genus of Lamiaceae. Brazilian J. Bot. 1–16.

Shang, X., Pan, H., Wang, X., He, H., Li, M., 2014. Leonurus japonicus Houtt.: ethnopharmacology, phytochemistry and pharmacology of an important traditional Chinese medicine. J. Ethnopharmacol. 152, 14–32. https://doi.org/10.1016/j.jep.2013.12.052.

Soares, A.C.F., Cabral, M.M.W., Martins, C.H.G., Ferreira, A.E., Bergamo, P.A.S., Omosa, L.K., Midiwo, J.O., Parreira, R.L.T., Heleno, V.C.G., 2017. Study of anti- tuberculosis activity behaviour of natural kaurane and trachylobane diterpenes compared with structural properties obtained by theoretical calculations. Nat. Prod. Commun. 12, 1934578X1701200 https://doi.org/10.1177/1934578X1701200521.

Soares, A., de, S., Pastore, J.F.B., Jardim, J.G., 2019. Lamiaceae no Rio Grande do norte. Brasil. Rodriguésia 70. https://doi.org/10.1590/2175-7860201970067.

Taamalli, A., Arraez-Roman, D., Abaza, L., Iswaldi, I., Fernandez-Gutierrez, A., Zarrouk, M., Segura-Carretero, A., 2015. LC-MS-based metabolite profiling of methanolic extracts from the medicinal and aromatic species Mentha pulegium and Origanum majorana. Phytochem. Anal. 26, 320–330.

Tada, M., Okuno, K., Chiba, K., Ohnishi, E., Yoshii, T., 1994. Antiviral diterpenes from Salvia officinalis. Phytochemistry 35, 539–541.

Taneja, N.K., Tyagi, J.S., 2007. Resazurin reduction assays for screening of anti- tubercular compounds against dormant and actively growing Mycobacterium tuberculosis, Mycobacterium bovis BCG and Mycobacterium smegmatis. J. Antimicrob. Chemother. 60, 288–293.

Thomsen, R., Christensen, M.H., 2006. MolDock: a new technique for high-accuracy molecular docking. J. Med. Chem. 49, 3315–3321. https://doi.org/10.1021/ jm051197e.

van Bondi, A., 1964. Van der Waals volumes and radii. J. Phys. Chem. 68, 441–451. Yao, J.-L., Fang, S.-M., Liu, R., Oppong, M., Liu, E.-W., Fan, G.-W., Zhang, H., 2016. A review on the terpenes from genus Vitex. Molecules 21, 1179. https://doi.org/ 10.3390/molecules21091179.

Zhang, H., Jin, B., Bu, J., Guo, J., Chen, T., Ma, Y., Tang, J., Cui, G., Huang, L., 2018. Transcriptomic insight into terpenoid biosynthesis and functional characterization of three diterpene synthases in Scutellaria barbata. Molecules 23, 2952. https://doi. org/10.3390/molecules23112952.

Zheng, C.-J., Lan, X.-P., Wang, Y., Huang, B.-K., Han, T., Zhang, Q.-Y., Qin, L.-P., 2012. A new labdane diterpene from Vitex negundo. Pharm. Biol. 50, 687–690. https:// doi.org/10.3109/13880209.2011.597410.

Capítulo 4

Neste capítulo, foi realizado o estudo de triagem virtual e caracterização química de diterpenos isolados as partes aéreas de *M. sidifolium*, além de avaliar o potencial farmacológico de compostos contra *Mycobacterium tuberculosis* e *Mycobacterium smegmatis* na busca por candidatos a novos fármacos. Sendo assim, foi realizado o estudo fitoquímico dos metabólitos com observação de dados espectrais de RMN uni e bidimensionais; avaliação *in silico* com construção de modelo preditivo e análise de citotoxicidade; e testes *in vitro* para susceptibilidade a microbactérias.

Por conseguinte, este trabalho resultou no isolamento de quatro componentes vegetais de *Mesosphaerum sidifolium*, o Pomiferin D, Salviol, Pomiferin E e 2α-hidroxisugiol, além de dois compostos fenólicos, ácido rosmarínico e ácido cafeico. Eles foram listados pela primeira vez no gênero *Mesosphaerum* e os sinais de ^{13}C NMR foram atribuídos de forma inequívoca. Assim, este trabalho teve como objetivo realizar a triagem virtual dos diterpenos da subtribo Hyptidinae por meio da análise fitoquímica da espécie *M. sidifolium* e, posteriormente, por meio de estudos de triagem virtual, selecionar os mais promissores com atividade antiturbecular frente a *M. tuberculosis*. Assim, contribuindo para o planejamento de compostos derivados de produtos naturais candidatos a novos fármacos, mais eficientes e seguros, para o tratamento da tuberculose auxiliado por ferramentas computacionais. O estudo *in-silico* evidenciou que dos 68 diterpenos da subtribo Hyptidinae analisados, 48 diterpenos tem provável atividade biológica contra *M. tuberculosis*. Os diterpenos isolados foram testados *in vitro* contra *M. tuberculosis* demonstrando MIC = 125 µM para 4 e 1, enquanto 2 e 3 -MIC = 250 µM. Esses compostos não apresentaram atividade biológica nessas concentrações para *M. smegmatis*.

Four diterpenes identified *in silico* were isolated from Hyptidinae and demonstrated *in vitro* activity Against *Mycobacterium tuberculosis*

Andreza Barbosa Silva Cavalcanti[a], Mayara dos Santos Maia[a], Pedro Thiago Ramalho de Figueiredo[a], Alex France Messias Monteiro[a], Renata Priscila Barros de Menezes[a], Roseana Araújo Ramos Meireles[a], Gabriela Cristina Soares Rodrigues[a], Ana Rita Rodrigues de Almeida Silva[a], Jociano da Silva Lins[a], Laísa Vilar Cordeiro[a], Valnês S. Rodrigues Junior[b,c], Ana Paula O. T. Castelo Branco[c], Maria de Fátima Agra[c]; Luciana Scotti[a], Marcelo Sobral da Silva[a], Vicente Carlos de Oliveira Costa[a], Josean Fechine Tavares[a] and Marcus Tullius Scotti [a]*

[a]Program of Natural and Synthetic Bioactive Products (PgPNSB), Health Sciences Center, Federal University of Paraíba, João Pessoa, Brazil
[b]National Institute of Science and Technology in Tuberculosis (INCT-TB), Brazil
[c]Program of Biotechnology, Center for Biotechnology, Federal University of Paraíba, João Pessoa, Brazil.

*Corresponding author: mtscotti@gmail.com; Tel.: +55-83-998690415

Artigo publicado na revista: **Natural Product Research**
DOI: https://doi.org/10.1080/14786419.2022.2096604
Fator de impacto: 2.861

Abstract

Plants of Hyptidinae subtribe (Lamiaceae – family), as *Mesosphaerum sidifolium*, are a source of bioactive molecules. In the search for new drug candidates, we perform chemical characterization of diterpenes isolated from the aerial parts of *M. sidifolium* was carried out with uni- and bidimensional NMR spectral data, and evaluate *in silico* through the construction of a predictive model followed by *in vitro* testing *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. Resulted in the isolation of four components: Pomiferin D (1), Salviol (2), Pomiferin E (3) and 2α-hydroxysugiol (4), as well as two phenolic compounds, rosmarinic and caffeic acids. *In silico* model identified 48 diterpenes likely to have biological activity against *M. tuberculosis*. The diterpenes isolated were tested *in vitro* against *M. tuberculosis* demonstrating MIC = 125 µM for 4 and 1, while 2 and 3 -MIC = 250 µM. These compounds did not show biological activity at these concentrations for *M. smegmatis*.

KEYWORDS

Hyptidinae; diterpenes; *Mesosphaerum sidifolium*; *in silico* model; *Mycobacterium tuberculosis*

1. Introduction

The medicinal uses of plants have been documented throughout history (Casanova and Costa 2017; Koparde et al. 2019). Brazil has one of the largest vegetation covers in the world, containing six biomes and more than 12,000 species of plants. The innate abundance of natural products in Brazil puts it in a privileged position to explore active therapeutics (Barreiro and Bolzani 2009; Saraiva et al. 2015; Berlinck et al. 2017). To identify active metabolites, we can perform virtual screenings of related compound libraries to elucidate robust candidates with desired properties (Alves et al. 2017).

The Lamiaceae family is composed of 295 genera and more than 7000 species of cosmopolitan distributed herbs and shrubs (Cavalcanti et al. 2019; Mesquita et al. 2019). In Brazil, 46 genus and 524 species of this family are present (Soares et al.

2019). Lamiaceae species metabolites are known worldwide for their various uses in ethnopharmacology and nutraceuticals, but more studies are needed to characterize the active phytochemicals. (Frezza et al. 2019). Diterpenes are abundant secondary metabolites in the Lamiaceae family (Zhang et al. 2018). The Hyptidinae clade contains many abietane diterpenes, known for high rates of endemism, and is well distributed in Brazil (Bridi et al. 2021).

Among the species of Hyptidinae to be studied, *Mesosphaerum sidifolium* (L'Hérit.), Harley & J.F.B. Pastore = *Hyptis umbrosa*, Salzm. ex Benth. is at the forefront of interest. Popularly known in Brazil as ‘bamburral’ and ‘Aleluia de serrote’, this species is a perennial sub-shrub (da Silva and Andrade 2004; Harley and Pastore 2012). *Mesosphaerum* plants were commonly used in folk medicine to treat headaches, menstrual cramps, the flu, gout, and digestive problems, as well as being known as an antimicrobial and antiseptic (Basílio et al. 2007). While *M. sidifolium* (MS) leaves are used to treat nasal and auricular disorders, they also have expectorant, carminative and tonic functions (Agra et al. 2008). Previous studies have shown that MS-ethanolic extract has potent antitumor activity (Rolim et al. 2017), as well as analgesic and anti-inflammatory activity (Dos Anjos et al. 2017; Scotti et al. 2018).

Studies of other species of Lamiaceae showed potential activity against *M. tuberculosis* of some compounds obtained such as phenolic derivatives, diterpenes, flavonoids, phenyl propanoids and steroids (Siddiqui et al. 2012; Lirio et al. 2014; Cavalcanti et al. 2021).

In light of the antimicrobial activity of *Mesosphaerum* species, these extracts and similar compounds serve as promising leads to fight infections such as tuberculosis (TB). T B is an infectious disease caused primarily by *M. tuberculosis*, which affects the respiratory tract (Basílio et al. 2007; Yu et al. 2020). It was diagnosed in more than 10 million people in 2016 and is responsible for 1.45 million deaths per year. Therapeutic regimens are available to treat TB in diagnosed patients. Those with active TB should seek treatment immediately and be sure to finish their treatment regimen. Unfortunately, inadequate treatment schemes have allowed for the development and spread of resistant strains of *M. tuberculosis* (Giacobbo et al. 2017; Muradás et al. 2018). Multiple studies have shown that natural products, particularly

diterpenes, display significant biological activity against TB compared to second line TB therapeutics (Matos et al. 2015; Silva et al. 2017; Soares et al. 2017).

The aim of our study is to use natural compounds to derive new, safer, and more efficient drug candidates to treat TB by using computational tools. Thus, this work performs a virtual screening of the Hyptidinae diterpenes through phytochemical analysis of the *M. sidifolium* species. That way, using the results of these screenings, we select the most promising compounds for antitubercular activity and follow with *in vitro* tests against *M. tuberculosis*.

2. Results and discussion

2.1. Phytochemistry

Six compounds were isolated from the aerial parts of *M. sidifolium* (Supplemental material Figure S1), four of which (**1–4**) are abietane diterpenes first reported in this genus. Spectral data for compounds 1–3 are found in the supplemental material (Tables S1, S2 and S3, respectively). Highlighting the reassignment of chemical shifts of ^{13}C, as Pomiferin D (Supplementary material Figures S2–S7), Pomiferin E and 2α-hydroxysugiol (Tables S1, S3 and S4, respectively).

Table 1. Activity of compounds against *M. tuberculosis* H37Ra and *M. smegmatis* mc^{2}155.

Compounds	**MIC (μM)**[a]	
	Mycobacterium tuberculosis	*Mycobacterium smegmatis*
INH[b]	3.8	-
RIF[b]	0.02	-
MOX[b]	0.2	0.2
2α-hydroxysugiol	125	>500
Salviol	250	>500
Pomiferim E	250	>500
Pomiferim D	125	>500

[a] MIC values reported here were observed in two independent experiments or the highest value found among three independent tests.
[b] INH, isoniazid; RIF, rifampicin; MOX, moxifloxacin.

Compound **4** was isolated in an amorphous green powder. The infrared spectrum (IR) demonstrated absorption at 3408 cm^{-1}, characteristic of the presence of phenolic hydroxyl. This which is corroborated by absorption at 1269 cm^{-1}, referring to the C–O attached to an aromatic ring (Supplementary material Figure S8). The ^1H

NMR (Nuclear magnetic resonance) spectrum showed two singlets in the region of aromatic protons with chemical shifts at δ_H 7.88 and 6.72 ppm, attributed to H-14 and H-11, respectively. Furthermore, a doublet at δ_H 1.24 ppm (J = 7.0 Hz) plus a septet at δ_H 3.18 ppm were observed and recognized as the isopropyl unit. These signals are compatible with the abietane diterpene skeleton and were attributed to H-14, H-11, 3H-16, 3H-17 and H-15, respectively (Supplementary material Table S4) (Gonzalez et al. 1989). Three singlets at δ_H 0.97, 1.01 and 1.22 ppm were observed, which, by comparison with the literature, were attributed to the methyls CH_3-18, CH_3-19 and CH_3-20 (Supplementary material Table S4) (Gonzalez et al. 1989). There was also a triplet of triplets at δ_H 4.09 ppm, attributed to a methoxy hydrogen. The ^{13}C NMR spectrum of Compound 4 showed 19 signals, corresponding to 20 carbon atoms. Of these, seven signals were attributed to nonhydrogenated carbon, five signals were attributed to methinic carbons, three signals were attributed to methylene carbons and four signals were attributed to methyl carbons. The presence of only two methinic sp^2 carbons implies oxygenation at C-12 and a protective effect explains the chemical shift at δ_C 109.45 ppm (C-11). This proposal is corroborated by the chemical shift at δ_C 26.75 ppm, which was attributed to C-15, under protection γ of the OH group at C-12 (Supplementary material Figure S10). The signals of δ_H 1.01 ppm (CH_3-19) and δ_C 32.56, 34.80, 48.75 and 50.33 ppm, were connected to C-18, C-4, C-5 and C-3, respectively. Furthermore, δ_H 0.97 ppm (CH_3-18) and the signals δ_C 22.29, 34.80, 48.75 and 50.33 ppm, were identified as C-19, C-4, C-5 and C-3, respectively. The C-7 signal was reinforced by the correlation of δ_H 7.88 ppm (H-14) with δ_C 197.87 ppm (C-7) (Supplementary material Figures S11 and Figures S14 and S15).

2.2. QSAR modelling

The predictive model showed excellent performance, with a ROC curve of 0.95 for external testing and 0.95 for internal cross-validation (Supplementary material Figure S16). The Matthews correlation coefficient (MCC) was 0.80 in the external test, revealing a robust model, in addition to excellent indices of sensitivity, specificity, accuracy and precision (Supplementary material Table S5).

We found that 48 of the 68 diterpenes analyzed displayed probable activity against *M. tuberculosis*, as shown in supplementary material Table S6. We also

observed that most species in this subtribe occur in regions of Brazil. The four isolated diterpenes were classified as active against *M. tuberculosis* with sufficient confidence (Supplementary material Table S7).

2.3. Prediction of absorption, distribution, excretion and toxicity properties

Additionally, the number of violations of Lipinski's rule (LIP) were calculated for each compound. This parameter indicates whether a molecule has good oral bioavailability (absorption or permeability), determined by four physicochemical parameters: molecular mass, octanol/water partition coefficient, number of hydrogen bonding donor atoms and number of hydrogen bonding atoms (Barreiro and Fraga 2014; Piccirillo and do Amaral 2018). Lipinski concluded that if a molecule violates two or more parameters, they are very likely to be poorly permeable (Capecchi et al. 2019). As seen in Supplementary Material Table S8, the maximum number of violations found was one parameter in the molecules Salviol and Pomiferin D, indicating that the four diterpenes have a high probability of being orally bioavailable.

The risk of cytotoxicity was also assessed and revealed that the four diterpenes did not present risks of cytotoxicity in the evaluated parameters. The oral absorption rate (topological surface area) was calculated as values of % ABS > 89%, suggesting this as a highly available route (Supplementary material Table S8). With these analyses, the diterpenes isolated from *M. sidifolium* indicate activity against the bacterium *M. tuberculosis* and excellent statistics relating to bioavailability, oral absorption, and cytotoxicity. Thus, these diterpenes are compounds that may find success in treating TB, and should be examined further.

2.4. Molecular docking

To identify a possible mechanism of action of these diterpenes in *M. tuberculosis*, a study of molecular docking was carried out. In docking, the interactions and energies of interaction between the molecule and target active site are analyzed and compared with the inhibitors of these proteins and drugs that are already used in

the treatment of the disease. An aminotransferase (PDB ID 5YHV) and an oxidoreductase (5VRN) and their respective inhibitors were obtained from the Protein Data Bank. The diterpenes had better interaction energy than some control drugs and had energy similar to or better than the inhibitors in the two proteins, as can be seen in supplementary material Figure S17.

The drug terizidone showed the best coupling energy in the two proteins, aminotransferase, and oxidoreductase (InhA). For the aminotransferase enzyme, the diterpenes had lower binding energy than the cocrystallized inhibitor, but Pomiferin D stands out, as the nest diterpene binding energy (–96.08 kJ.mol^{-1}) and presented energy like the cocrystallized inhibitor (–96.70 kJ.mol^{-1}). However, all diterpenes showed better interaction energies than the prothionamide, ethionamide, cycloserine and pyrazinamide controls (Supplementary material Figure S17). Pomiferin D showed interactions with residues similar to terizidone control and/or complexed inhibitor; among the hydrogen bonds it is possible to highlight Ser100, Arg242 and Ser101, in addition to the steric interactions with the residues Ser100, Tyr125, Asn174, Lys234 and Tyr128. Some of these residues mentioned are also found in the compounds 2α-hydroxysugiol, Salviol and Pomiferin E, and we hypothesize that these similarities contribute to the good anchoring energies presented by all the analysed compounds, reinforcing the possible antitubercular potential and mechanism of action of the diterpenes of this study in relation to interaction with aminotransferase (Supplementary material Figure S18 and Table S9).

For the InhA enzyme, Pomiferin E (–99.30 kJ.mol^{-1}) presented the best coupling value compared with the other diterpenes; however, all diterpenes analysed showed better energies than the controls, except for terizidone (–108.71 kJ.mol^{-1}) (Supplementary material Figure S19). Pomiferin E did not show hydrogen bonds with the protein, but the steric bonds gave a better coupling to this compound than the other diterpenes. The steric interaction with the Leu218 residue in the protein is present in terizidone and all the diterpenes analyzed, however, it is not found in the prothionamide control (–56.43 kJ.mol^{-1}) or the complex inhibitor (Supplementary material Figure S19 and Table S9).

We analyzed of the possible metabolites of these diterpenes, their toxicities in humans and the ADME properties. Only metabolite 35 of Salviol (Supplementary material Table S10) presented a very high risk of toxicity on the reproductive system and the liver, with a score above 85 (Supplementary material Figures S20 and S21). All compounds showed high probability to be absorbed in the gastrointestinal tract and all were predicted like inhibitor the CYP2D6 (Supplementary material Table S11).

2.5. Antitubercular biological activity

After The molecules had their biological activities tested *in vitro* against *M. tuberculosis* H37Ra, whose data can be found in Table 1. The compounds 2α-hydroxysugiol and Pomiferin D showed greater activity against *M. tuberculosis* with Minimum Inhibitory Concentration (MIC) 125 μM, while Salviol and Pomiferin E showed MIC 250 μM. Thus, the result of the biological test validated the antitubercular activity prediction model used, in which all the diterpenes tested were promising against *M. tuberculosis*, corroborating the *in silico* study. However, these compounds did not at these concentrations show biological activity for *M. smegmatis*.

Studies reported by Soares et al. demonstrate that some kaurane and trachylobane diterpenes have antituberculosis activity, with MIC values similar than the second-choice antituberculosis drug Cycloserine (122.4–498 μM) (Soares et al. 2017). Thus, we suggest that the diterpenes identified in this study also have similar potency to Cycloserine.

In Lamiaceae, the diterpenes 6,7-dehydroroyleanone, 6β,7α-dihydroxyroyleanone, 7α-acetoxy-6β-hydroxyroyleanone, horminone, coleon U quinone and *ent*-3β-acetoxy-kaur-15-en-17-oic acid showed activity evaluated and proven against *M. tuberculosis*, thus, this work corroborates that the species of Lamiaceae are promising sources of diterpenes, important ethnomedicinal agents against tuberculosis (Baldin et al. 2018; Cavalcanti et al. 2021; Ndjoubi et al. 2021).

3. Experimental

3.1. Instrumentation and general procedures

The detailed description is provided in the supplementary material.

3.2. Plant material

The detailed description is provided in the supplementary material.

3.3. Isolation of chemical constituents

The detailed description is provided in the supplementary material.

3.4. Quantitative structure-activity relationship (QSAR) modelling

The detailed description is provided in the supplementary material.

3.5. Predicting absorption, distribution, excretion, toxicity properties and hepatic metabolites and ADME approach

The detailed description is provided in the supplementary material.

3.6. Molecular docking

The detailed description is provided in the supplementary material.

3.7. Mycobacterial susceptibility assay

The detailed description is provided in the supplementary material.

4. Conclusions

We performed phytochemical characterization of the *M. sidifolium* plant through the isolation of Pomiferin D, Salviol, Pomiferin E, 2α-hydroxysugiol, rosmarinic acid and caffeic acid. Isolated abietane diterpenes are herein reported for the first time in the *Mesosphaerum* genus, and the ^{13}C signal assignment was done unambiguously, for Pomiferin D, Pomiferin E and 2α-hydroxysugiol. *In silico* analysis that includes virtual screening of antitubercular activity and ADMET prediction suggested that 48 of

the 68 Hyptidinae diterpenes were likely to be active against *M. tuberculosis*. Of the four diterpenes analyzed in this research, Salviol showed the highest potential for activity accompanied by high risk for reproductive and/or renal toxicity, which is causing high concern regarding its use. However, all four diterpenes were tested in vitro against *M. tuberculosis* and demonstrated antitubercular activity at very high concentrations: 2α-hydroxysugiol and Pomiferin D with MIC = 125 μM, while Salviol and Pomiferin E showed MIC = 250 μM. As predicted, these compounds did not show biological activity at these concentrations for *M. smegmatis*.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq), grant number 309648/2019-0 and 431254/2018-4 and by the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

Acknowledgments

The present work was carried out with the support the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq), grant numbers 309648/2019-0, 431254/2018-4, and 404440/2016-9 (to V.S.R.J.), and by the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

References

Agra MdF, Silva KN, Basílio IJLD, Freitas PFd, Barbosa-Filho JM. 2008. Survey of medicinal plants used in the region Northeast of Brazil. Revista brasileira de farmacognosia.18:472-508.

Alves VM, Braga RC, Muratov EN, Andrade CH. 2018. Quimioinformática: uma introdução. Química Nova.41:202-212.

Baer-Dubowska W, Latacz G. 2018. Session 13. Metabolic Activation of Xenobiotics-Therapeutic Target and Drug Monitoring. Acta Biochimica Polonica.65:78-81.

Baldina VP, Scodro RBdL, Lopes-Ortiz MA, Almeida AL, Gazimd ZC, Ferarrese L, Faiões VdS, Torres-Santos EC, Piresa CTA, Caleffi-Ferracioli KR, Siqueira VLD, Cortez DG, Cardoso RF. 2018. Anti-*Mycobacterium tuberculosis* activity of essential oil and 6,7-dehydroroyleanone isolated from leaves of *Tetradenia riparia* (Hochst.) Codd (Lamiaceae). Phytomedicine. 47:34-39.

Barreiro EJ, Bolzani VdS. 2009. Biodiversidade: fonte potencial para a descoberta de fármacos. Química nova.679-688.

Barreiro EJ, Fraga CAM. 2014. Química Medicinal-: As bases moleculares da ação dos fármacos: Artmed Editora.

Basílio I, de FAgra M, Rocha EA, Leal CKA, Abrantes HF. 2007. Estudo farmacobotânico comparativo das folhas de *Hyptis pectinata* (L.) Poit. e *Hyptis suaveolens* (L.) Poit (Lamiaceae). acta farmacéutica bonaerense.25:518.

Berlinck RG, Borges WdS, Scotti MT, Vieira PC. 2017. A química de produtos naturais do Brasil do século XXI. Química Nova.40:706-710.

Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Thiel K, Wiswedel B. 2009. KNIME-the Konstanz information miner: version 2.0 and beyond. AcM SIGKDD explorations Newsletter.11:26-31.

Bridi H, de Carvalho Meirelles G, von Poser GL. 2020. Subtribe Hyptidinae: a promising source of bioactive metabolites. Journal of ethnopharmacology.113225.

Capecchi A, Awale M, Probst D, Reymond JL. 2019. PubChem and CHEMBL beyond Lipinski. Molecular informatics.38:1900016.

Casanova LM, Costa SS. 2017. Synergistic Interactions in Natural Products: Therapeutic Potential and Challenges. REVISTA VIRTUAL DE QUIMICA.9:575-595.

Cavalcanti ABS, Barros RPC, Costa VCdO, Sobral da Silva M, Tavares JF, Scotti L, Scotti MT. 2019. Computer-aided chemotaxonomy and bioprospecting study of diterpenes of the Lamiaceae family. Molecules.24:3908.

Cavalcanti ABS, Figueiredo PTR, Veloso CAG, Rodrigues GCS, Maia MdS, Monteiro AFM, Rodrigues-Junior VS, Castelo-Branco APOT, Agra MdF, Braz-Filho R, Sobral da Silva M, Tavares JF, Costa VCdO, Scotti L, Scotti MT. 2021. A new labdane diterpene from the aerial segments of *Leptohyptis macrostachys* (L'Hérit.) Harley & J.F.B. Pastore. Phytochemistry Letters. 42:117-122.

Daina A, Michielin O, Zoete V. 2017. SwissADME: a free web tool to evaluate pharmacokinetics, druglikeness and medicinal chemistry friendliness of small molecules. Scientific Reports. 7:42717.

da Silva VA, Andrade LdHC. 2004. O significado cultural das espécies botânicas entre indígenas de Pernambuco: o caso Xucuru. Biotemas.17:79-94.

Dauzhenka T, Anishchenko I, Kundrotas P, Vakser I. 2020. Protein Docking Refinement with Systematic Conformational Search-Application to Models Inside the Docking Funnel. Biophysical Journal.118:516a.

Dragon 7.0 (software for molecular descriptor calculation) 2019-12-18. Available from https://chm.kode-solutions.net/pf/dragon-7-0/

Dos Anjos KS, Araújo-Filho HG, Duarte MC, Costa VC, Tavares JF, Silva MS, Almeida JR, Souza NA, Rolim LA, Menezes IR. 2017. HPLC-DAD analysis, antinociceptive and anti-inflammatory properties of the ethanolic extract of *Hyptis umbrosa* in mice. EXCLI journal.16:14.

Frezza C, Venditti A, Serafini M, Bianco A. 2019. Phytochemistry, chemotaxonomy, ethnopharmacology, and nutraceutics of Lamiaceae. In: Studies in natural products chemistry. Elsevier. p. 125-178.

Giacobbo BC, Pissinate K, Rodrigues-Junior V, Villela AD, Grams ES, Abbadi BL, Subtil FT, Sperotto N, Trindade RV, Back DF. 2017. New insights into the SAR and drug combination synergy of 2-(quinolin-4-yloxy) acetamides against *Mycobacterium tuberculosis*. European journal of medicinal chemistry.126:491-501.

Gonzalez AG, Aguiar ZE, Luis JG, Ravelo AG, Domínguez XA. 1989. Minor quinone methide diterpenoids from the roots of *Salvia texana*. Journal of Natural Products.52:1231-1236.

Harley R, Pastore J. 2012. A generic revision and new combinations in the Hyptidinae (Lamiaceae), based on molecular and morphological evidence. Phytotaxa.58:1-55.

Koparde AA, Doijad RC, Magdum CS. 2019. Natural products in drug discovery. In: Pharmacognosy-Medicinal Plants. IntechOpen.

Kumari M, Tiwari N, Chandra S, Subbarao N. 2018. Comparative analysis of machine learning based QSAR models and molecular docking studies to screen potential anti-tubercular inhibitors against InhA of *Mycobacterium tuberculosis*. International Journal of Computational Biology and Drug Design.11:209-235.

Lirio SB, Macabeo APG, Paragas EM, Knorn M, Kohls P, Franzblau SG, Wang Y, Aguinaldo MAM. 2014. Antitubercular constituents from *Premna odorata* Blanco. 154:471-474.

Matos PM, Mahoney B, Chan Y, Day DP, Cabral MM, Martins CH, Santos RA, Bastos JK, Page PCB, Heleno VC. 2015. New non-toxic semi-synthetic derivatives from natural diterpenes displaying anti-tuberculosis activity. Molecules.20:18264-18278.

Mesquita LSSd, Luz TRSA, Mesquita JWCd, Coutinho DF, Amaral FMMd, Ribeiro MNdS, Malik S. 2019. Exploring the anticancer properties of essential oils from family Lamiaceae. Food Reviews International.35:105-131.

Monteiro AF, Scotti MT, Speck-Planche A, Barros RPC, Scotti L. 2020. *In Silico* Studies for Bacterystic Evaluation against Staphylococcus aureus of 2-Naphthoic Acid Analogues. Current topics in medicinal chemistry.20:293-304.

Muradás TC, Abbadi BL, Villela AD, Macchi FS, Bergo PF, de Freitas TF, Sperotto NDM, Timmers L, Norberto de Souza O, Picada JN, et al. 2018. Pre-clinical evaluation of quinoxaline-derived chalcones in tuberculosis. PLoS One.13:e0202568. Epub 2018/08/17.

Ndjoubi KO, Sharma R, Badmus JA, Jacobs A, Jordaan A, Marnewick J, Warner DF, Hussein AA. 2021 Antimycobacterial, Cytotoxic, and Antioxidant Activities of Abietane Diterpenoids Isolated from P*lectranthus Madagascariensis*. Plants. 10:175.

Piccirillo E, do Amaral AT. 2018. Virtual screening of bioactive compounds: concepts and aplications. Quimica Nova.41:662-677.

Rolim TL, Meireles DRP, Batista TM, De Sousa TKG, Mangueira VM, De Abrantes RA, Pita JCLR, Xavier AL, Costa VCO, Batista LM. 2017. Toxicity and antitumor potential of *Mesosphaerum sidifolium* (Lamiaceae) oil and fenchone, its major component. BMC complementary and alternative medicine.17:347.

Sander T, Freyss J, von Korff M, Rufener C. 2015. DataWarrior: an open-source program for chemistry aware data visualization and analysis. Journal of chemical information and modeling.55:460-473.

Saraiva SRGL, Saraiva HCC, de Oliveira-Júnior RG, Silva JC, Damasceno CMD, da Silva Almeida JRG, Amorim ELC. 2015. A implantação do programa de plantas medicinais e fitoterápicos no sistema público de saúde no brasil: uma revisão de literatura. Revista Interdisciplinar de Pesquisa e Inovação.1.

Saroj D, Biswal B. Full wwPDB X-ray Structure Validation Report i○.

Scotti MT, Herrera-Acevedo C, Oliveira TB, Costa RPO, Santos SYKdO, Rodrigues RP, Scotti L, Da-Costa FB. 2018. SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. Molecules.23:103.

Siddiqui BS, Bhatti HA, Begum S, Perwaiz S. 2012. Evaluation of the antimycobacterium Activity of the constituents from *Ocimum basilicum* against *Mycobacterium tuberculosis*. Journal of Ethnopharmacology. 144:220-222.

Sidrônio MGS, Castelo Branco APOT, Abbadi BL, Macchi F, Silveira M D, Lock GA, Costa TD, Araújo DM, Cibulski S, Bizarro CV, Machado P, LA, Rodrigues-Junior VS, 2021. Effects of tafenoquine against active, dormant and resistant *Mycobacterium tuberculosis*. Tuberculosis. 128: 102089.

Silva AN, Soares ACF, Cabral MM, de Andrade AR, Silva M, Martins CH, Veneziani R, Ambrósio SR, Bastos JK, Heleno VC. 2017. Antitubercular activity increase in labdane diterpenes from *Copaifera* oleoresin through structural modification. Journal of the Brazilian Chemical Society.28:1106-1112.

Soares AC, Cabral MM, Martins CH, Ferreira AE, Bergamo PA, Omosa LK, Midiwo JO, Parreira RL, Heleno VC. 2017. Study of Anti-Tuberculosis Activity Behaviour of Natural Kaurane and Trachylobane Diterpenes Compared with Structural Properties Obtained by Theoretical Calculations. Natural product communications.12:1934578X1701200521.

Soares AdS, Pastore JFB, Jardim JG. 2019. Lamiaceae no Rio Grande do Norte, Brasil. Rodriguésia.70.

Taneja NK, Tyagi JS. 2007. Resazurin reduction assays for screening of anti-tubercular compounds against dormant and actively growing *Mycobacterium tuberculosis, Mycobacterium bovis* BCG and *Mycobacterium smegmatis*. Journal of antimicrobial chemotherapy.60:288-293.

Ulubelen A, Topcu G. 1992. Abietane diterpenoids from *Salvia pomifera*. Phytochemistry.31:3949-3951.

Xia Y, Zhou Y, Carter DS, McNeil MB, Choi W, Halladay J, Berry PW, Mao W, Hernandez V, O'Malley T. 2018. Discovery of a cofactor-independent inhibitor of *Mycobacterium tuberculosis* InhA. Life science alliance.1.

Yu W-y, Lu P-X, Tan W-g. 2020. Tuberculosis Control in Migrating Population: Springer.

Zhang H, Jin B, Bu J, Guo J, Chen T, Ma Y, Tang J, Cui G, Huang L. 2018. Transcriptomic insight into terpenoid biosynthesis and functional characterization of three diterpene synthases in *Scutellaria barbata*. Molecules.23:2952.

Conclusão

CONSIDERAÇÕES FINAIS

Devido a existência de uma grande variedade de diterpenos na família Lamiaceae, foi construído um banco de dados e realizado uma análise quimiotaxonômica. Esse banco de dados com mais de 4115 moléculas está disponível na ferramenta web SistematX (https://sistematx.ufpb.br). O estudo quimiotaxonômico das subfamílias Lamiaceae, demonstrou altas taxas de precisão (>80%) corroborando com estudo anterior proposto por Li e colaboradores. Dessa forma, através desta análise foi possível prever a localização dos diterpenos em suas subfamílias e auxiliar na busca de metabólitos secundários com características estruturais específicas, como compostos com potencial atividade biológica.

Através da triagem virtual com a criação de modelos preditivos e estudos de modelagem molecular foi possível propor que alguns diterpenos isolados das espécies *Leptohyptis macrostachys* e *Mesosphaerum sidifolium* da família Lamiaceae apresentavam probabilidade de serem ativas frente a *M. tuberculosis.* Sendo assim, realizou-se a triagem *in vitro* que corroborou os resultados do estudo *in silico*.

Portanto, este estudo resultou na ampliação do conhecimento químico e farmacológico de espécies da família Lamiaceae.

Anexos

Anexo I

Produção Científica: Artigos completos publicados

COMPUTER-AIDED CHEMOTAXONOMY AND BIOPROSPECTING STUDY OF DITERPENES OF THE LAMIACEAE FAMILY

AUTORES: Andreza Barbosa Silva Cavalcanti, Renata Priscila Costa Barros, Vicente Carlos de Oliveira Costa, Marcelo Sobral da Silva, Josean Fechine Tavares, Luciana Scotti and Marcus Tullius Scotti

REVISTA: ***Molecules*****,** 2019**,** Volume 29, 3908. doi:10.3390/molecules24213908

Fator de impacto: 4.411

A NEW LABDANE DITERPENE FROM THE AERIAL SEGMENTS OF *LEPTOHYPTIS MACROSTACHYS* (L'HÉRIT.) HARLEY & J.F.B. PASTORE

AUTORES: Andreza B. S. Cavalcanti, Pedro T.R.de Figueiredo, Carlos A. G. Veloso, Gabriela C. S. Rodrigues, Mayara dos S. Maia, Alex France Messias Monteiro, Valnês S. Rodrigues Junior, Ana P.O.T. Castelo-Branco, Maria de F. Agra, Raimundo B. Filho, Marcelo S. da Silva, Josean F. Tavares, Vicente C.de O. Costa, Luciana Scotti, Marcus T. Scotti

REVISTA: ***Phytochemistry Letters*****,** Volume 43, June 2021, Pages 117-122. https://doi.org/10.1016/j.phytol.2021.03.022

Fator de impacto: 1.679

LIGAND-AND STRUCTURE-BASED VIRTUAL SCREENING OF LAMIACEAE DITERPENES WITH POTENTIAL ACTIVITY AGAINST A NOVEL CORONAVIRUS (2019-NCOV)

AUTORES: Gabriela Cristina Soares Rodrigues, Mayara Dos Santos Maia, Renata Priscila Barros de Menezes, Andreza Barbosa Silva Cavalcanti, Natália Ferreira de Sousa, Érika Paiva de Moura, Alex France Messias Monteiro, Luciana Scotti, Marcus Tullius Scotti

REVISTA: ***Curr Top Med Chem*****.** 2020;20(24):2126-2145. doi: 10.2174/1568026620666200716114546.

Fator de impacto: 3.295

CONSENSUS ANALYSES IN MOLECULAR DOCKING STUDIES APPLIED TO MEDICINAL CHEMISTRY

AUTORES: Mayara dos Santos Maia, Gabriela Cristina Soares Rodrigues, Andreza Barbosa Silva Cavalcanti, Luciana Scotti, Marcus Tullius Scotti.

REVISTA: ***Mini-Reviews in Medicinal Chemistry*****,** 2020;20(14):1322-1340, doi: 10.2174/1389557520666200204121129.

Fator de impacto: 3.862

IN SILICO STUDIES OF LAMIACEAE DITERPENES WITH BIOINSECTICIDE POTENTIAL AGAINST _APHIS GOSSYPII_ AND _DROSOPHILA MELANOGASTER_

AUTORES: Gabriela Cristina Soares Rodrigues, Mayara dos Santos Maia, Andreza Barbosa Cavalcanti, Marcus Tullius Scotti e Luciana Scotti.

REVISTA: ***Molecules*** 2021, *26*(3), 766; https://doi.org/10.3390/molecules26030766

Fator de impacto: 4.411

COMPUTER-ASSISTED DISCOVERY OF COMPOUNDS WITH INSECTICIDAL ACTIVITY AGAINST _MUSCA DOMESTICA_ AND _MYTHIMNA SEPARATA_

AUTORES: Gabriela Cristina Soares Rodrigues, Mayara dos Santos Maia, Andreza Barbosa Cavalcanti, Marcus Tullius Scotti e Luciana Scotti.

REVISTA: ***Molecules*** 2021, *26*(3), 766; https://doi.org/10.3390/molecules26030766

Fator de impacto: 4.411

FOUR DITERPENES IDENTIFIED _IN SILICO_ WERE ISOLATED FROM HYPTIDINAE AND DEMONSTRATED _IN VITRO_ ACTIVITY AGAINST _MYCOBACTERIUM TUBERCULOSIS_

AUTORES: Andreza Barbosa Silva Cavalcanti, Mayara dos Santos Maia, Pedro Thiago Ramalho de Figueiredo, Alex France Messias Monteiro, Renata Priscila Barros de Menezes, Roseana Araújo Ramos Meireles, Gabriela Cristina Soares Rodrigues, Ana Rita Rodrigues de Almeida Silva, Jociano da Silva Lins, Laísa Vilar Cordeiro, Valnês S. Rodrigues Junior, Ana Paula O. T. Castelo Branco, Maria de Fátima Agra; Luciana Scotti, Marcelo Sobral da Silva, Vicente Carlos de Oliveira Costa, Josean Fechine Tavares and Marcus Tullius Scotti. .

REVISTA: ***Natural Product Research*** 2022; https://doi.org/10.1080/14786419.2022.2096604

Fator de impacto: 2.861

Anexo II

Material suplementar do artigo: A new labdane diterpene from the aerial segments of *Leptohyptis macrostachys* (L'Hérit.) Harley & J.F.B. Pastore

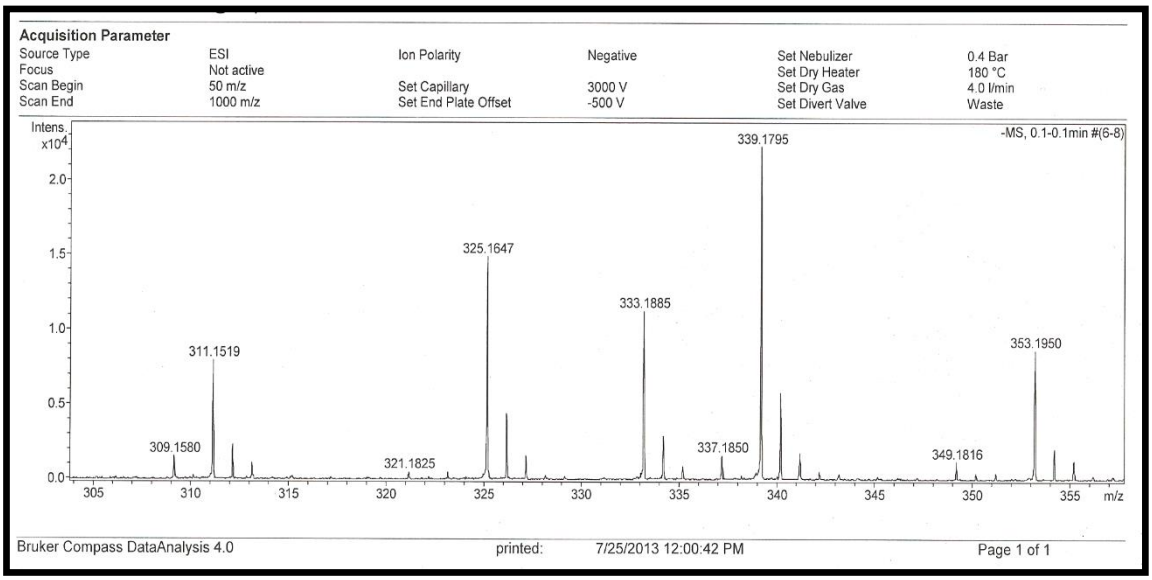


Figure S1. HR-ESI-MS spectrum of the compound **1**

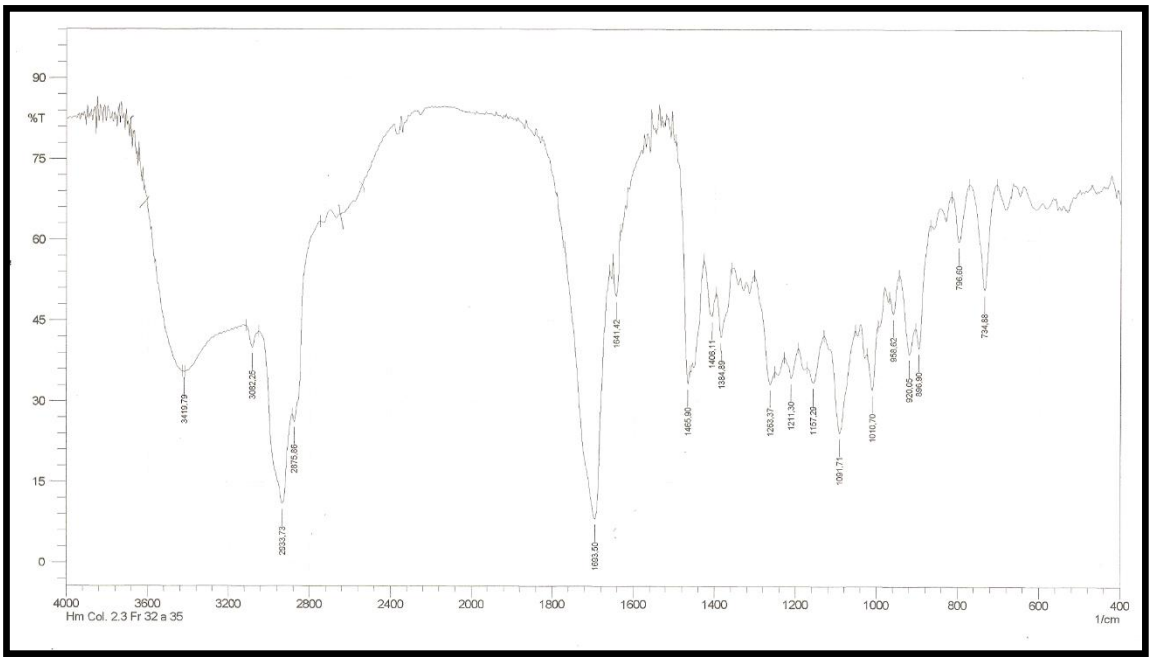


Figure S2. IR spectrum (in KBr) of the compound **1**

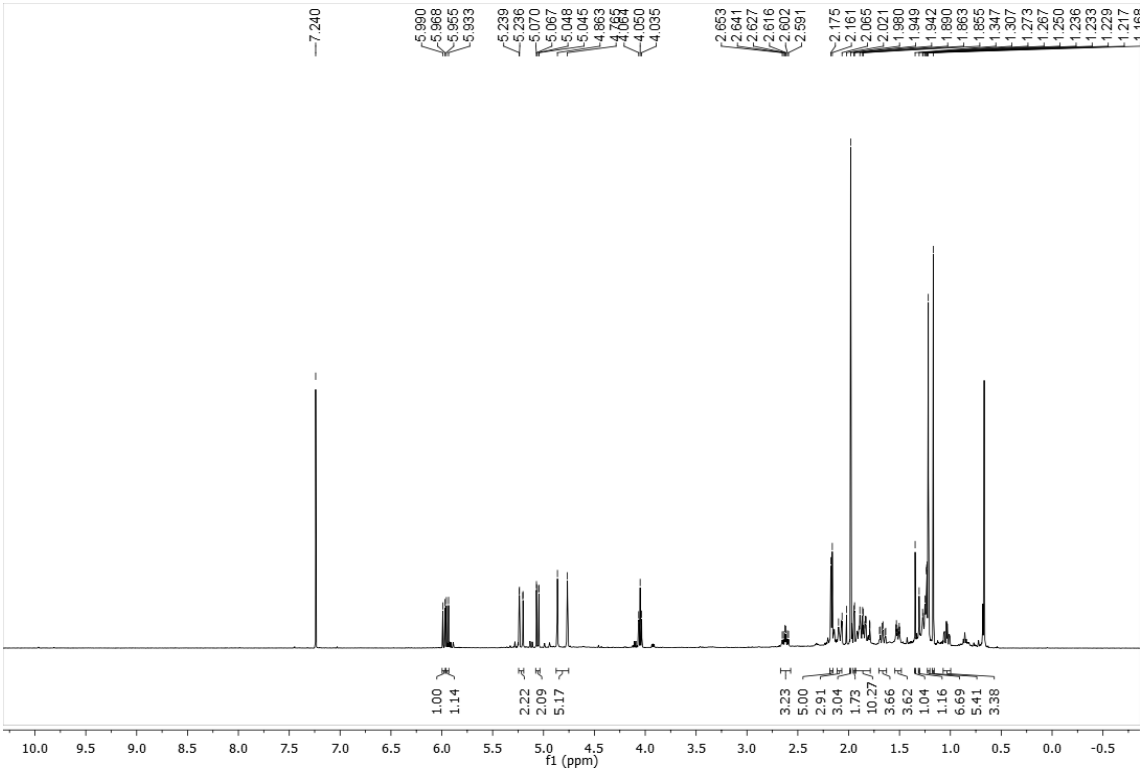


Figure S3. ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

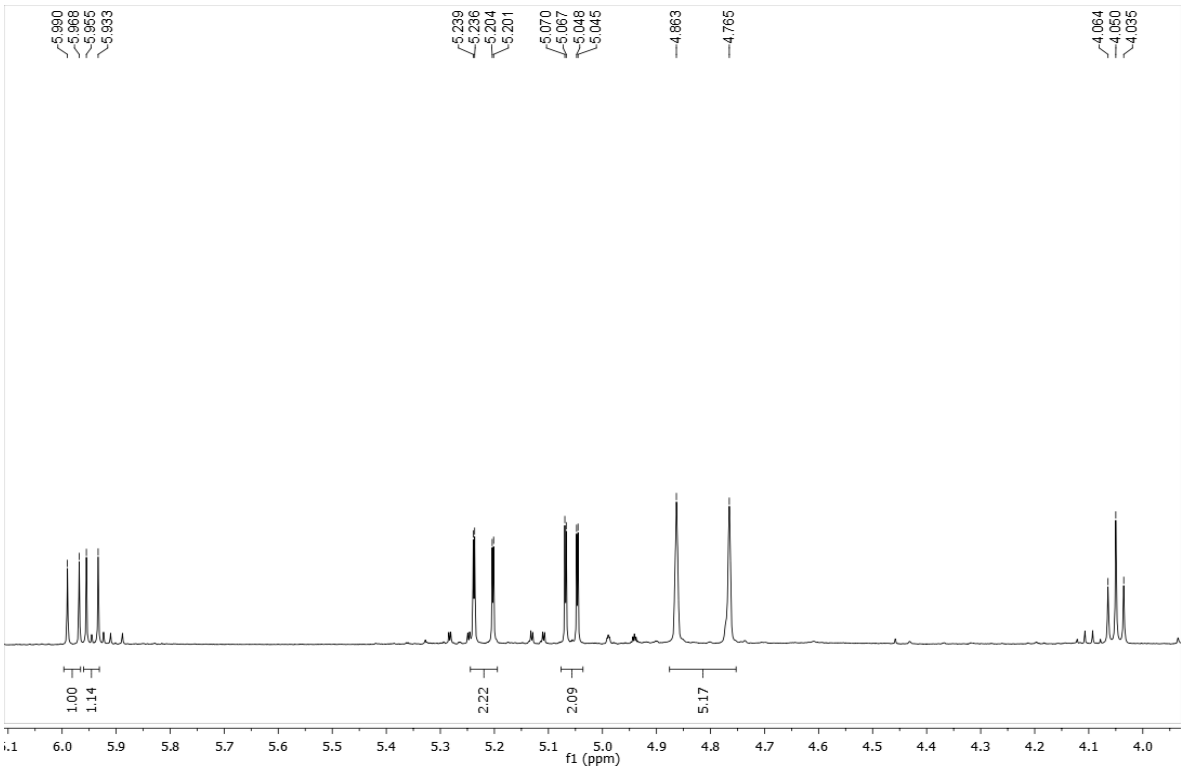


Figure S4. ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

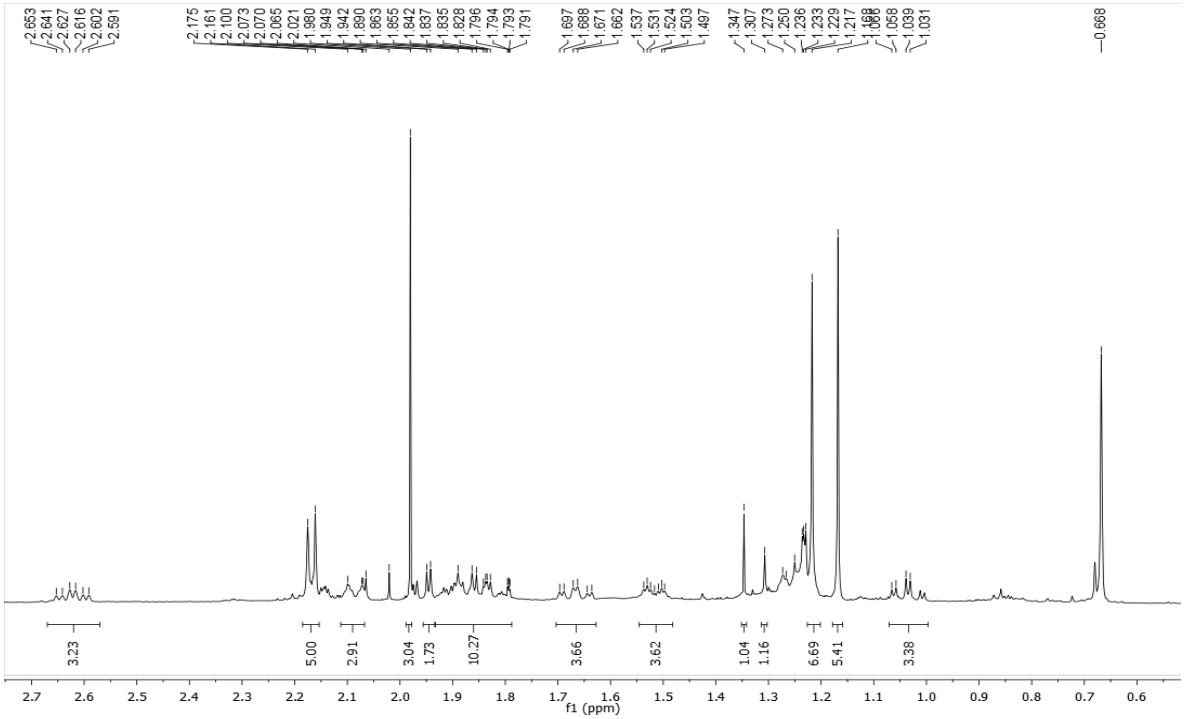


Figure S5. ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

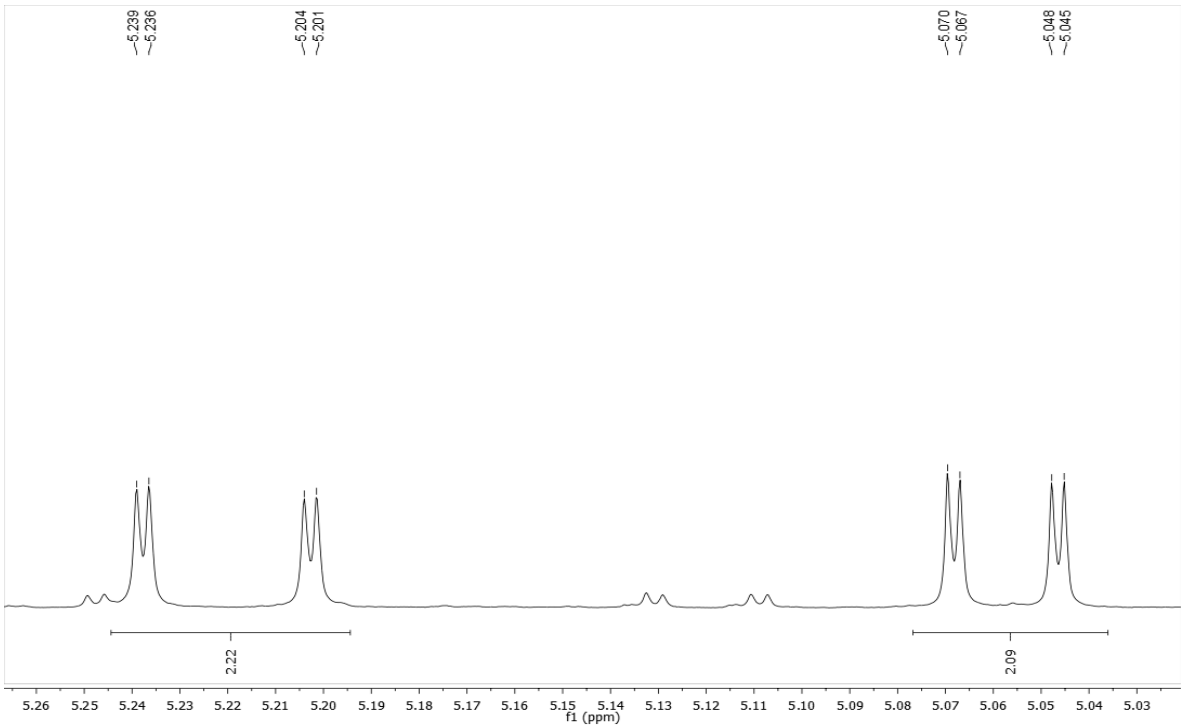


Figure S6. ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

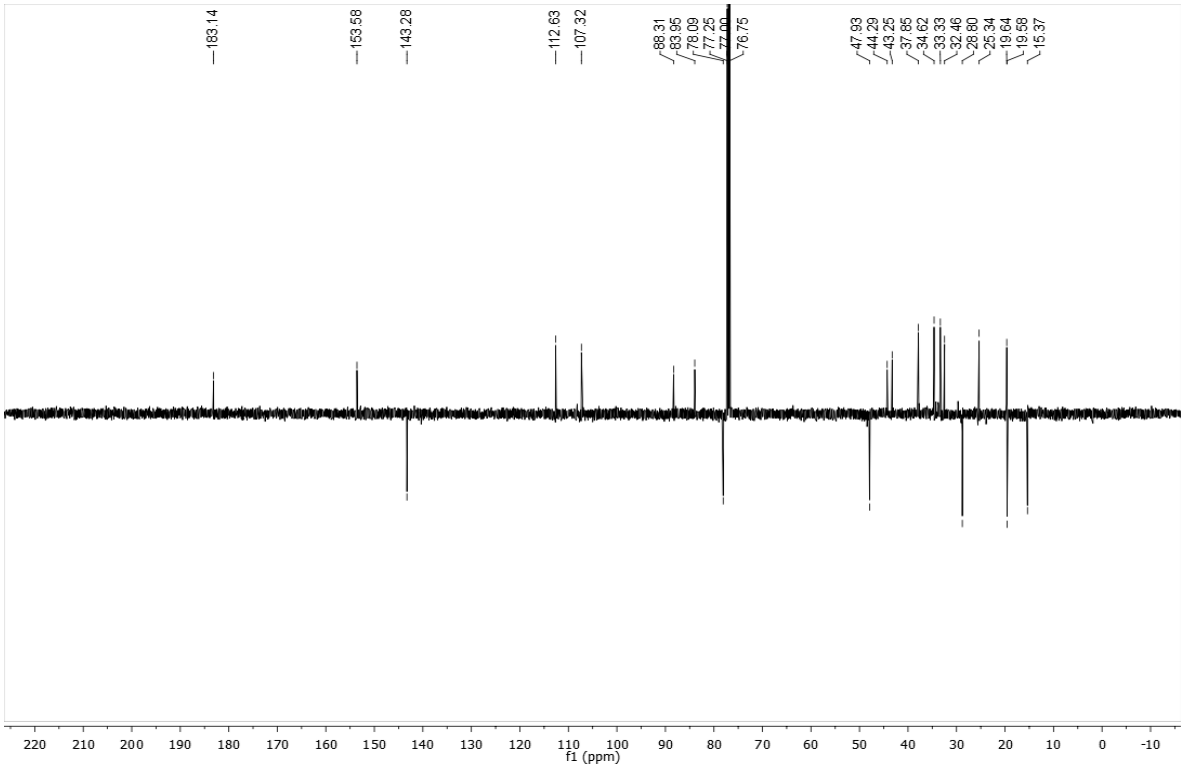


Figure S7. ^{13}C NMR-APT spectrum (125 MHz, $CDCl_3$) of the compound **1**

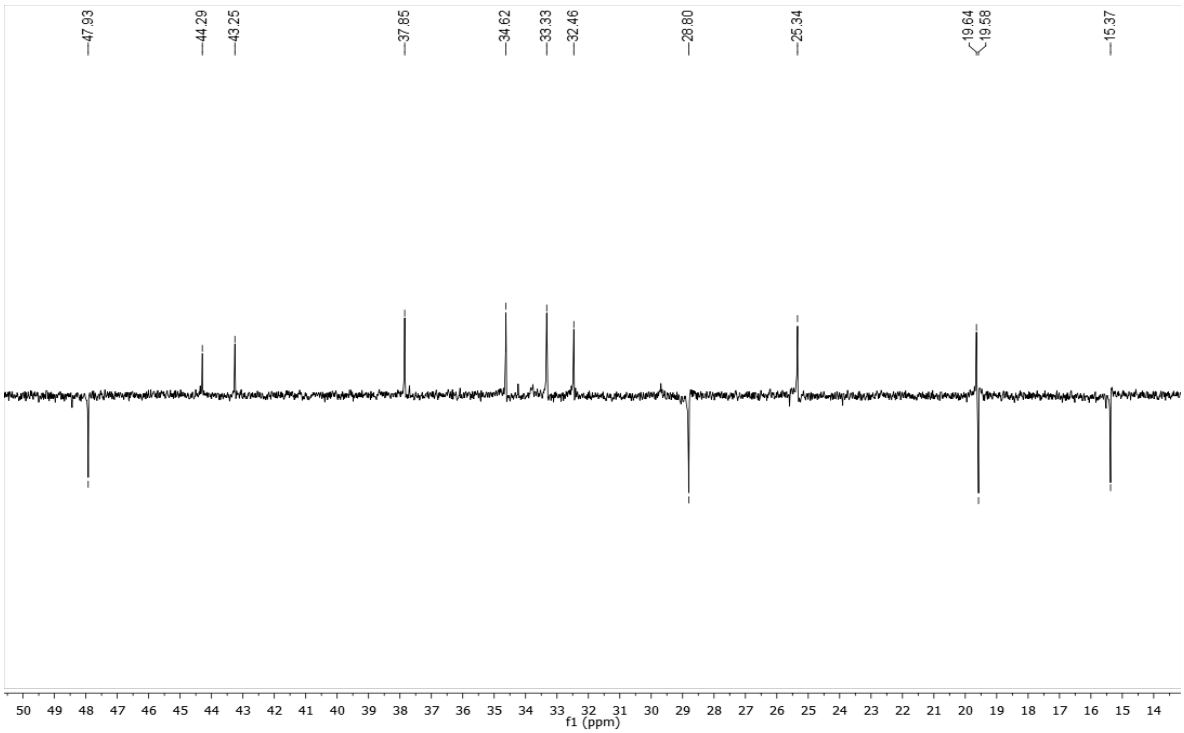


Figure S8. ^{13}C NMR-APT spectrum (125 MHz, $CDCl_3$) of the compound **1**

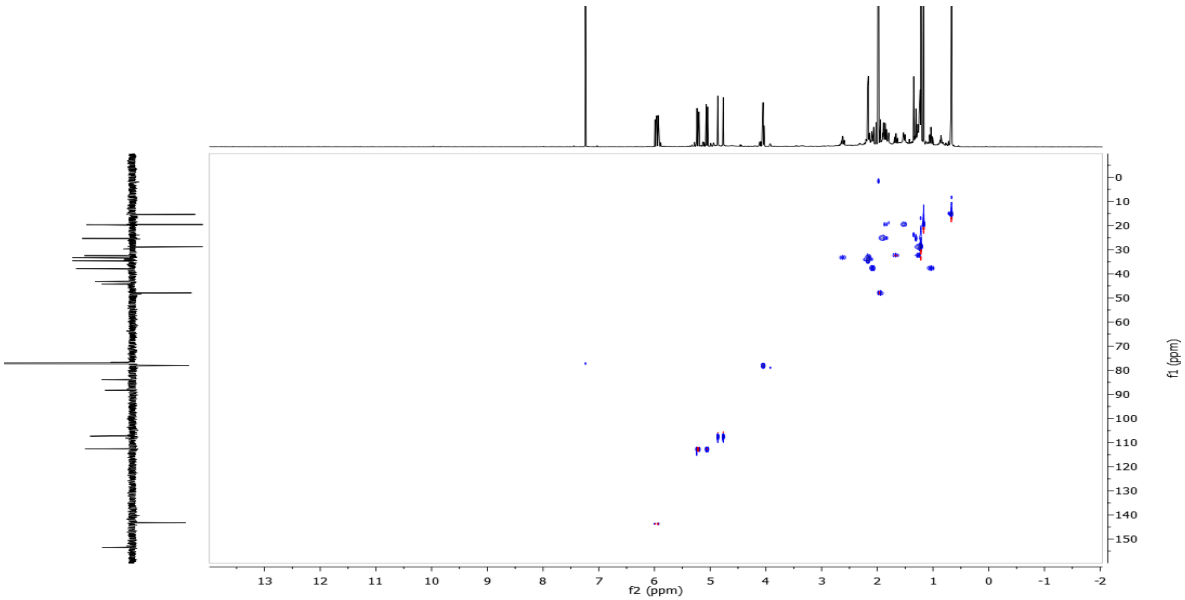


Figure S9. HMQC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

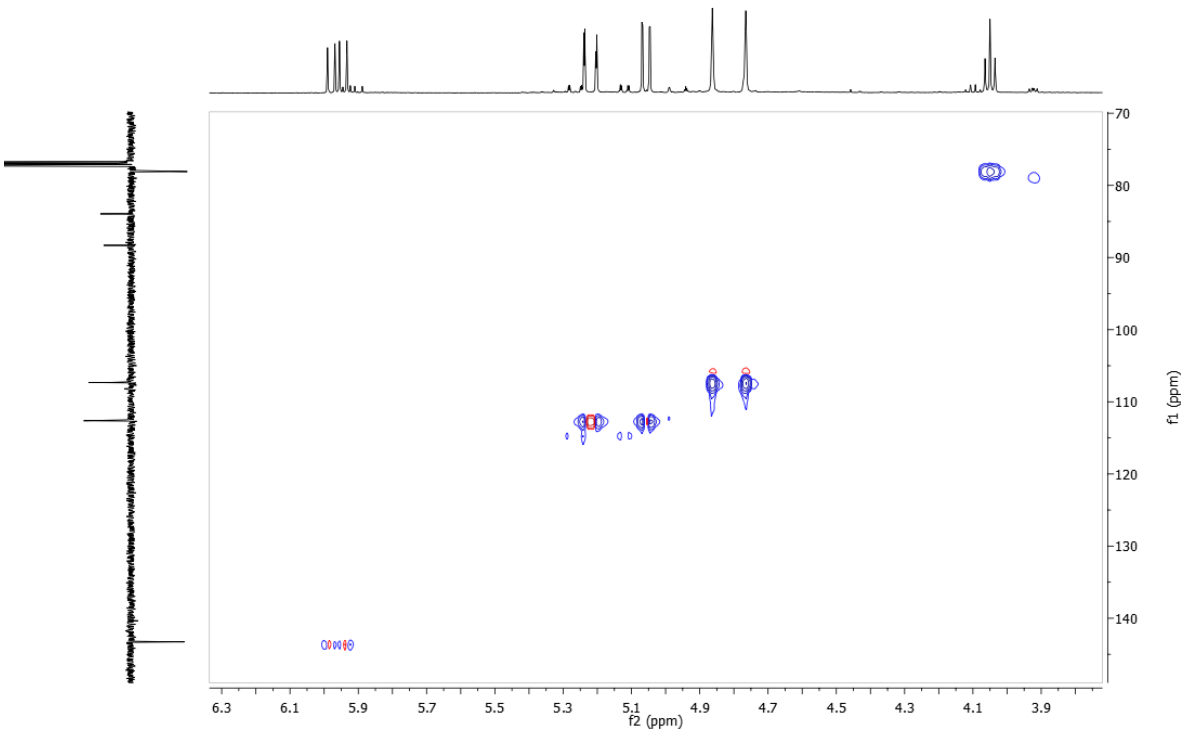


Figure S10. HMQC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

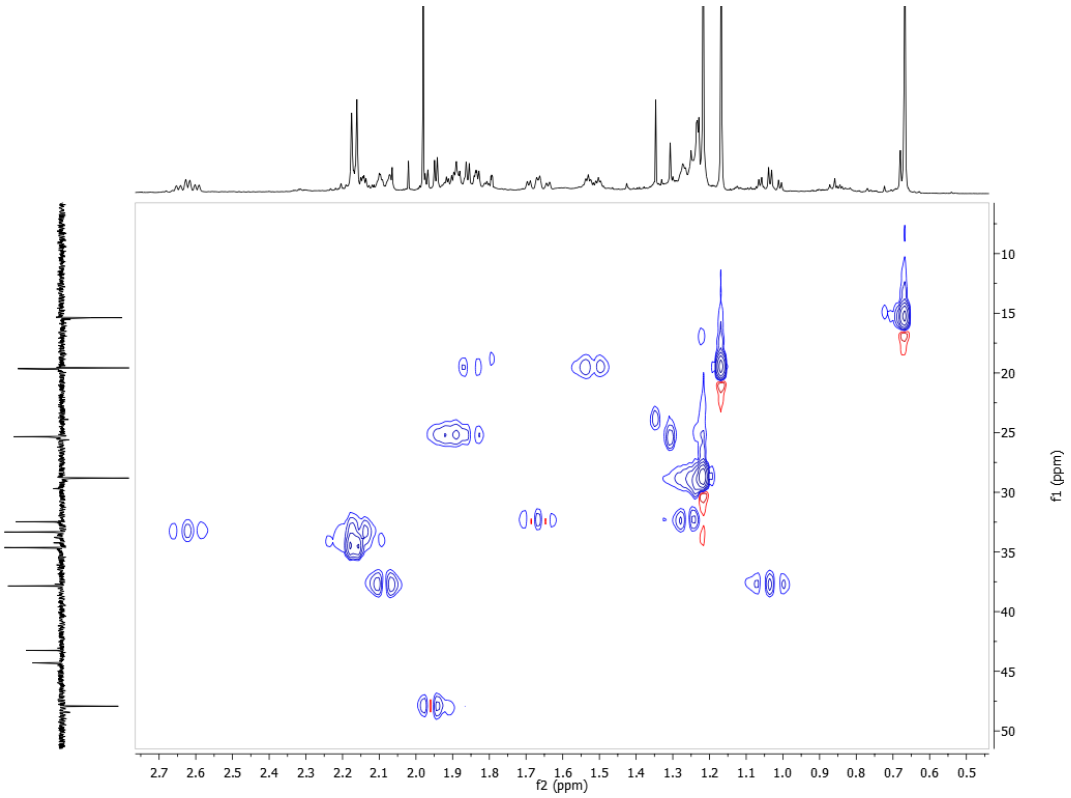


Figure S11. HMQC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

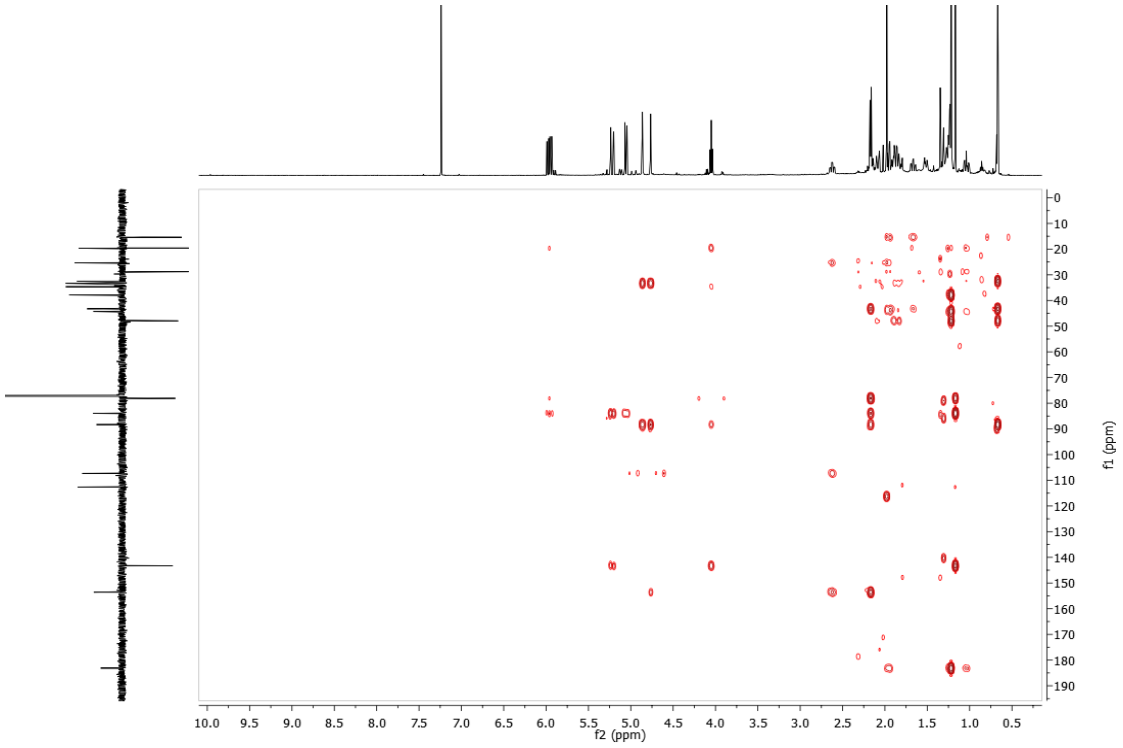


Figure S12. HMBC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

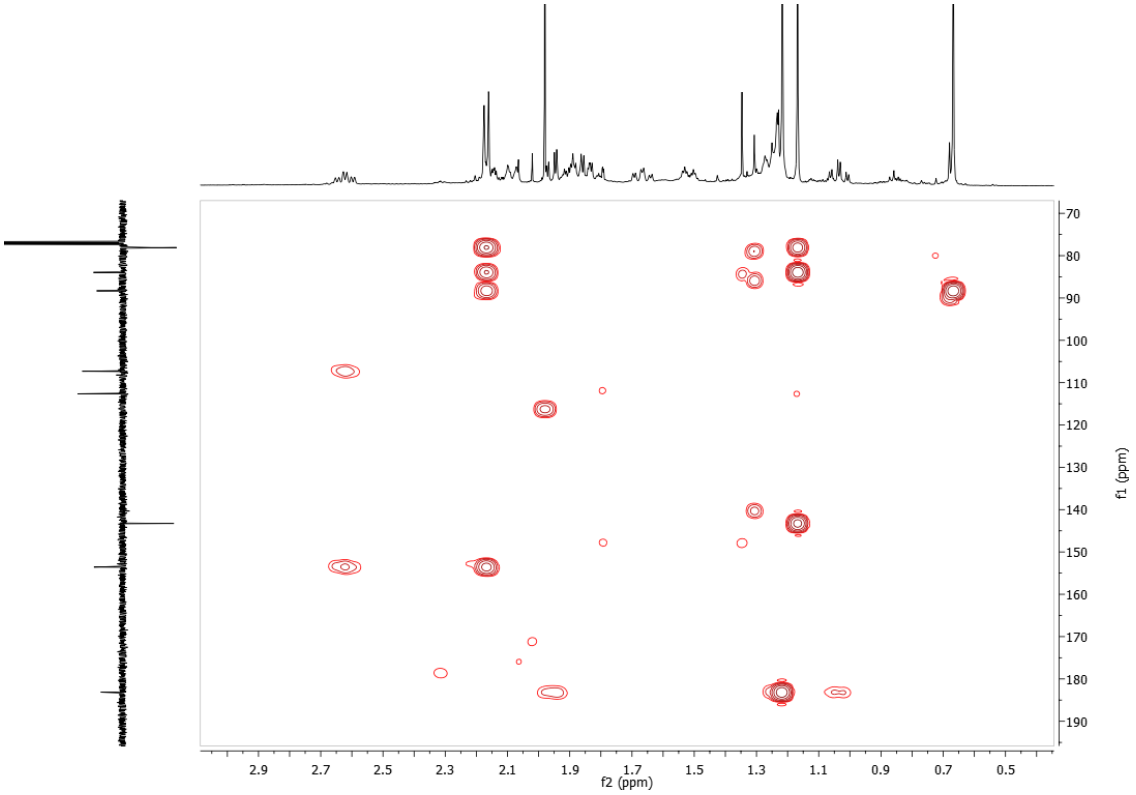


Figure S13. HMBC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

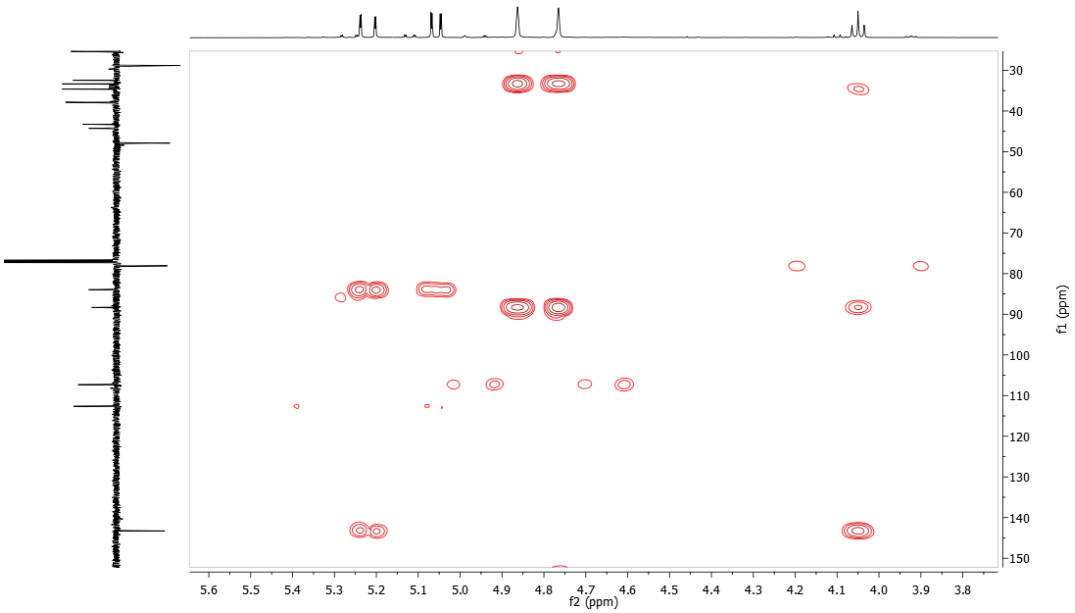


Figure S14. HMBC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

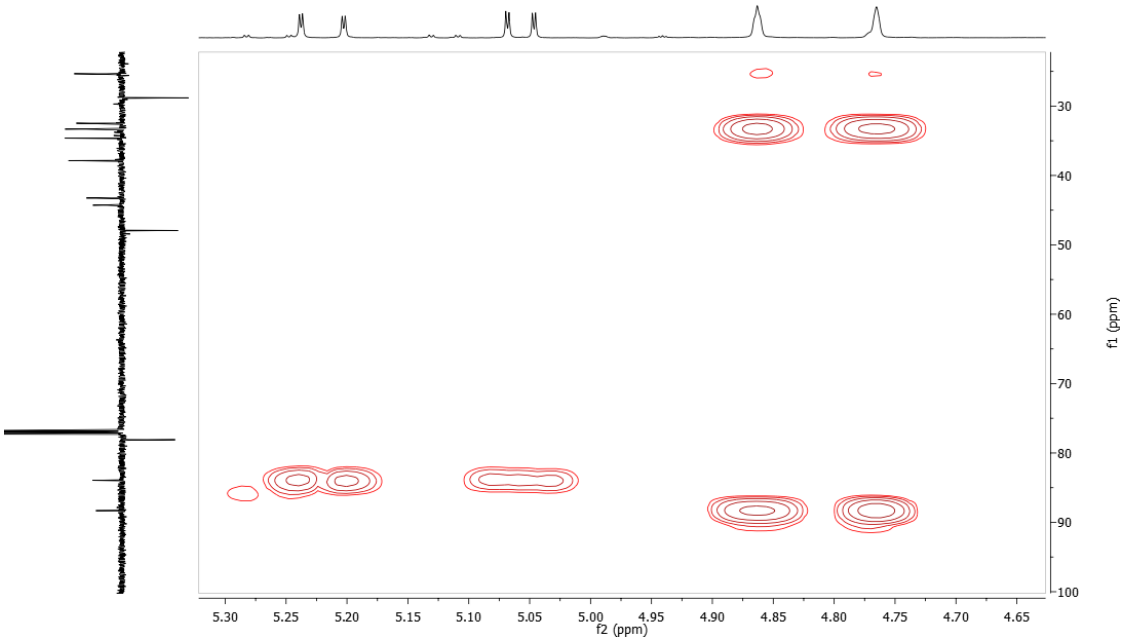


Figure S15. HMBC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

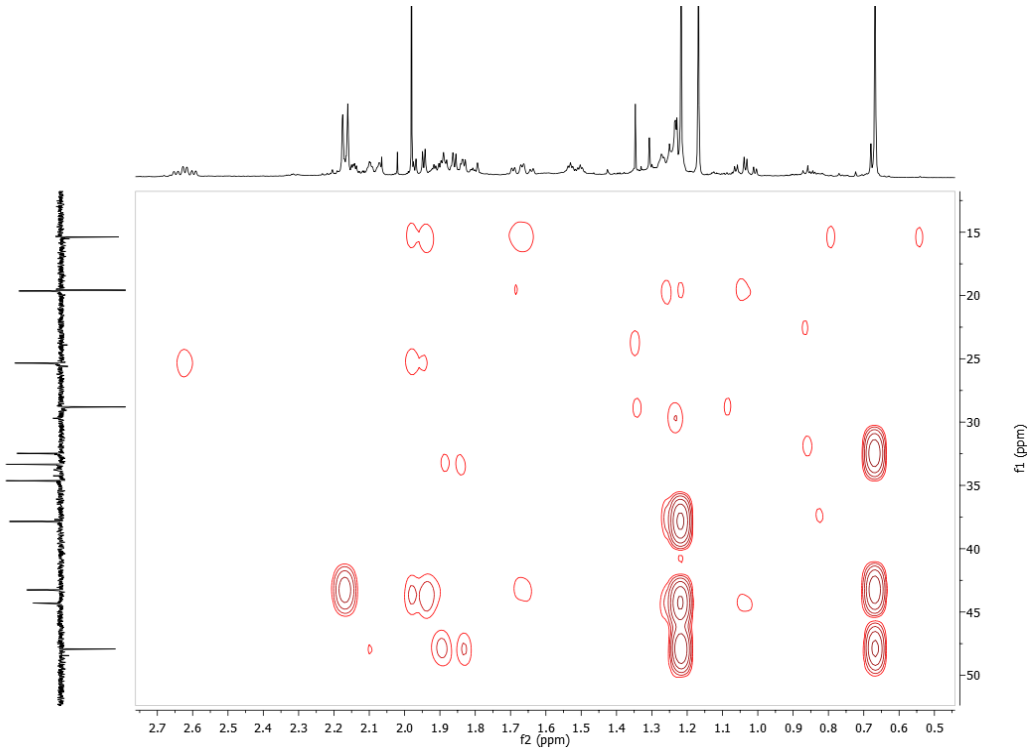


Figure S16. HMBC NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

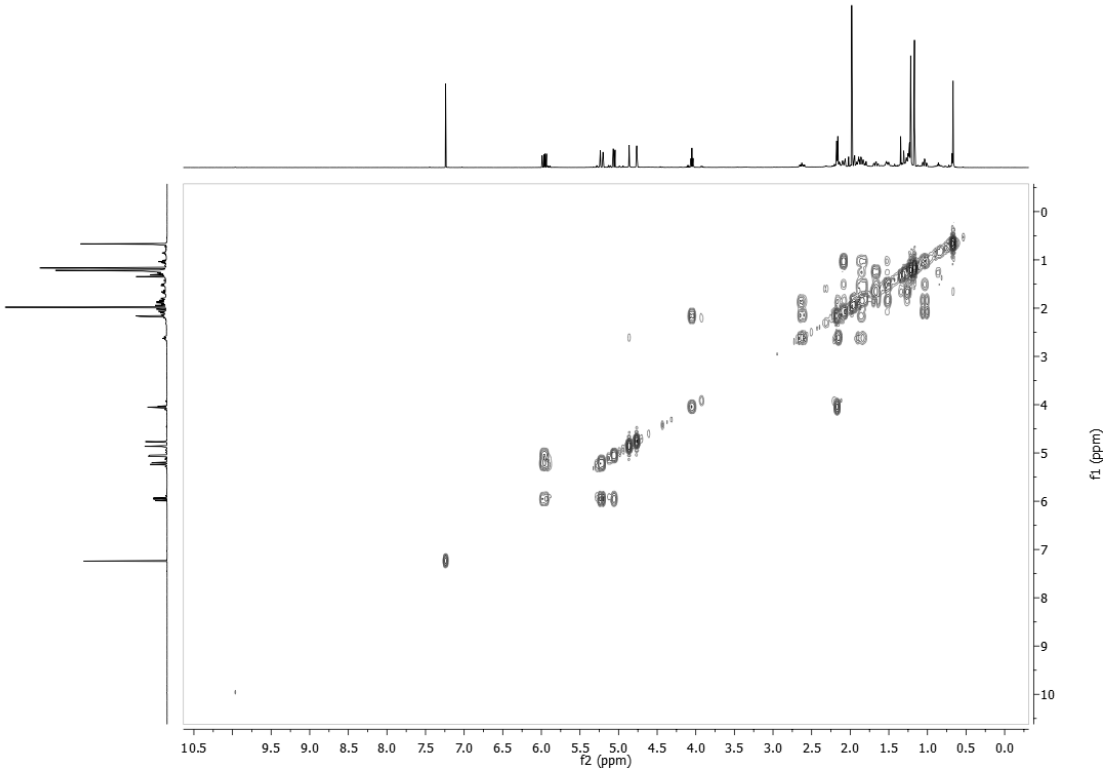


Figure S17. COSY NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

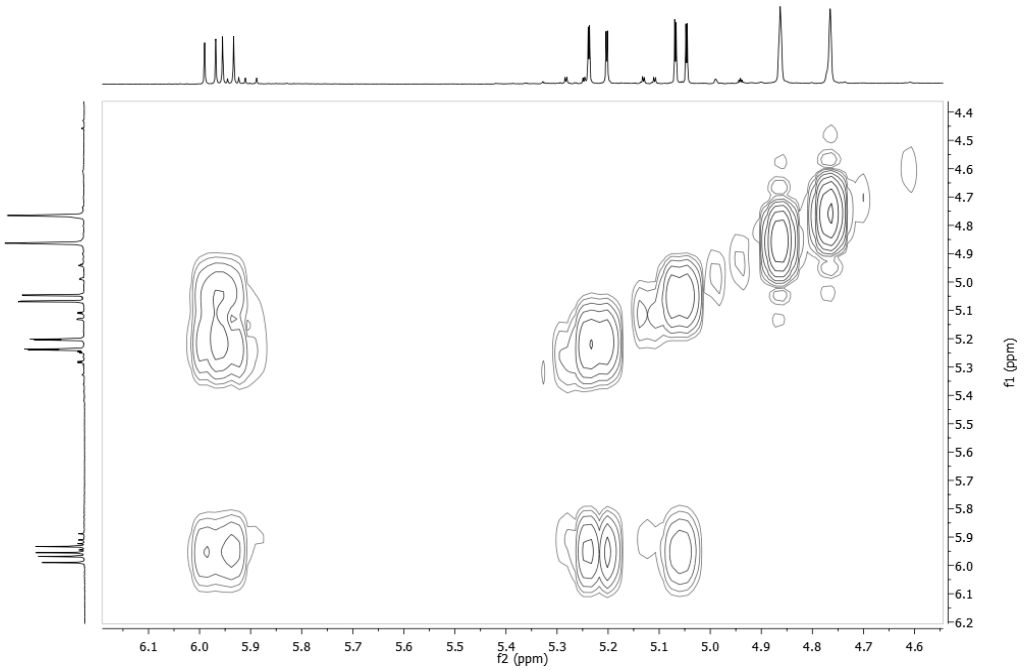


Figure S18. COSY NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

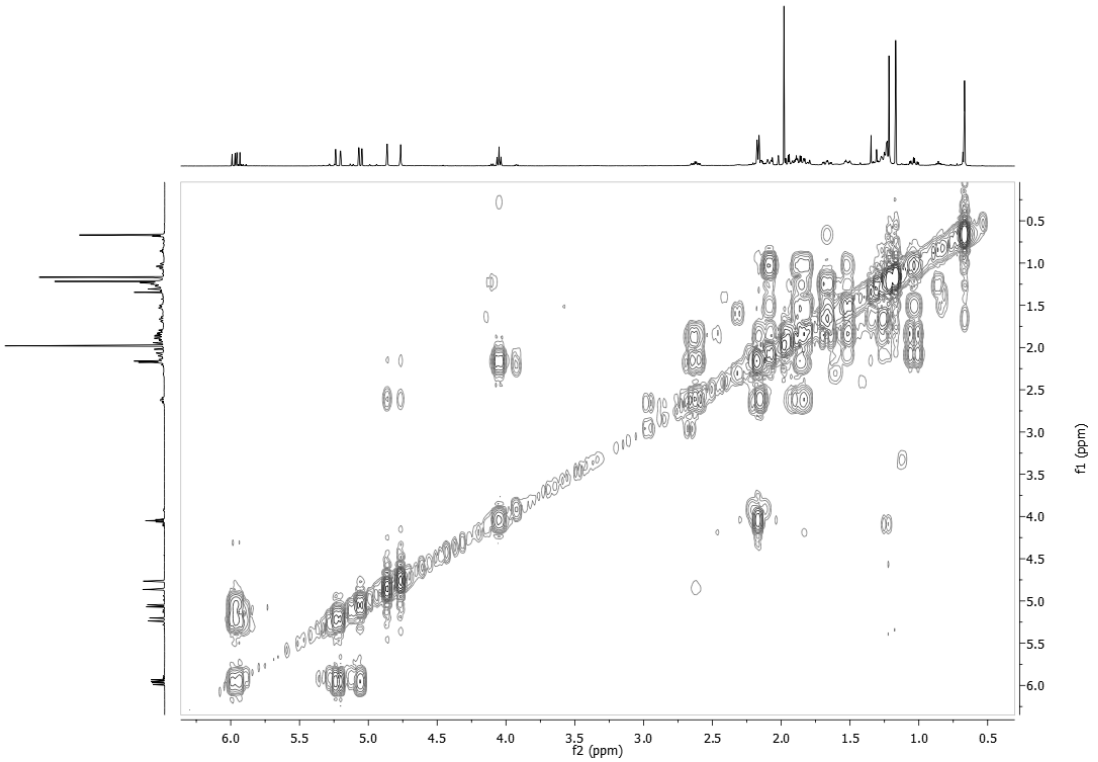


Figure S19. COSY NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

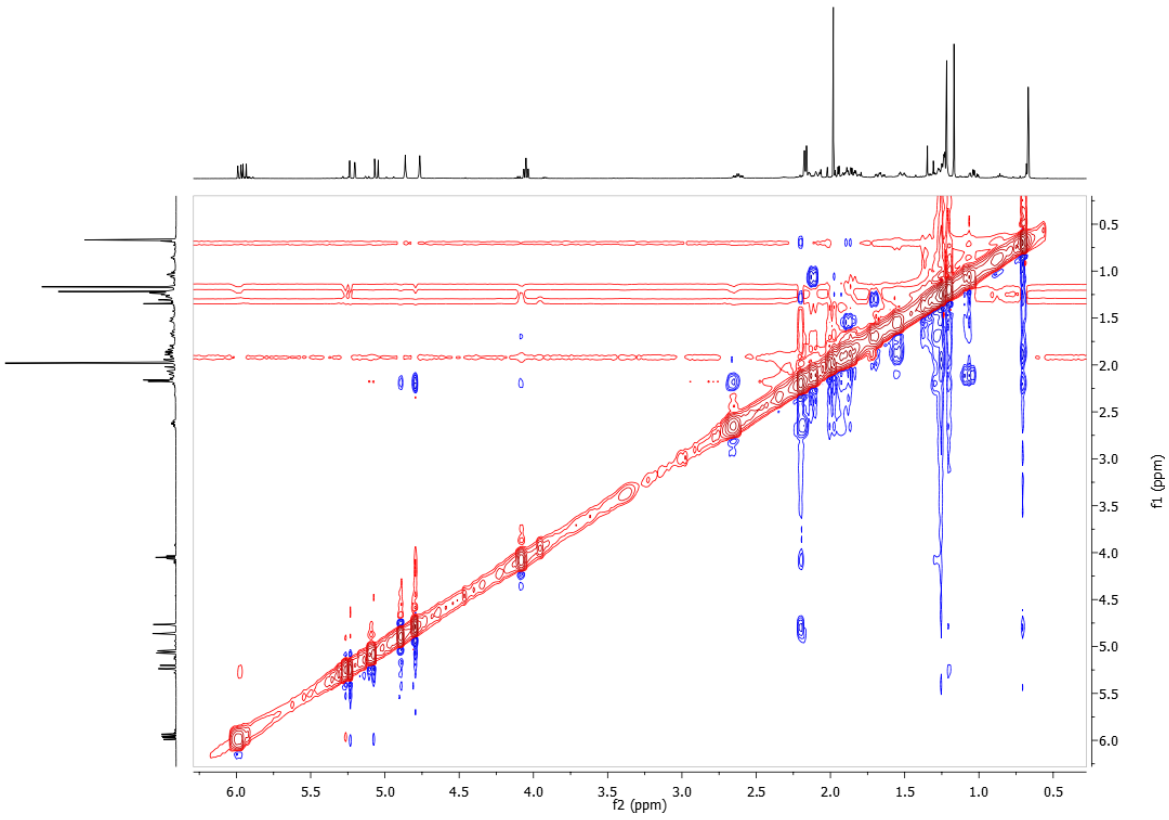


Figure S20. NOESY NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

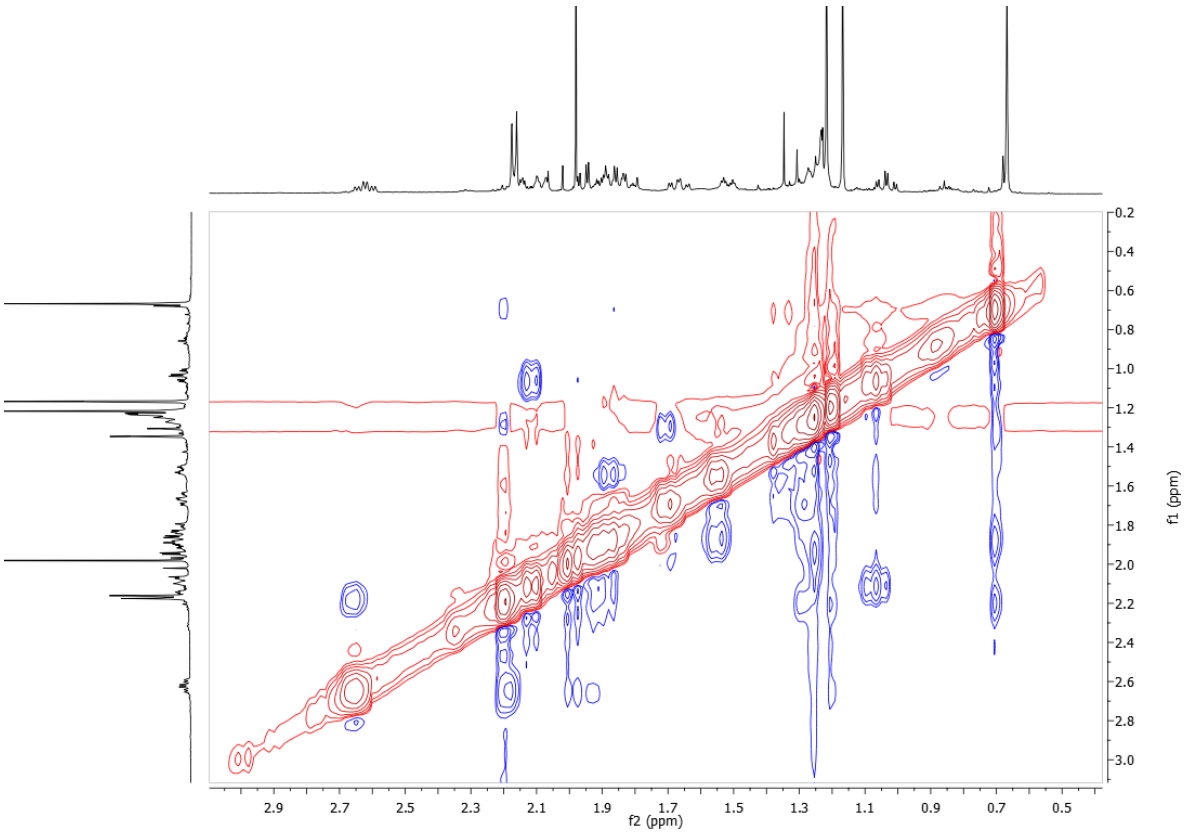


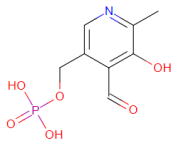
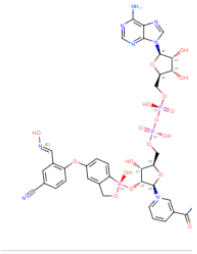
Figure S21. NOESY NMR spectrum (500 MHz, $CDCl_3$) of the compound **1**

Table S1. Comparison of calculated DFT-NMR of **1** and your isomer 12α-hydroxy-13α-methyl **1a**.

Position	δ_C (1)	δ_H (1)	δ_C (1a)	δ_H (1a)
1	32.5	1.91 (H1 ax), 1.21 (H1 eq)	32.4	2.11 (H1 ax), 1.53 (H1 eq)
2	21.1	1.45 (H2 ax), 2.26 (H2 eq)	21.1	1.53 (H2 ax), 2.28 (H2 eq)
3	37.6	1.21 (H3 ax), 2.11 (H3 eq)	37.5	1.20 (H3 ax), 2.12 (H3 eq)
4	43.3	-	43.5	-
5	49.6	2.37	50.2	2.17
6	25.9	1.88 (H6 ax), 2.25 (H6 eq)	25.8	1.85, (H6 ax), 2.29 (H6 eq)
7	34.4	2.93 (H7 ax), 2.07 (H7 eq)	34.0	2.83 (H7 ax), 2.00 (H7 eq)
8	154.3	-	153.4	-
9	88.4	-	88.4	-

10	45.4	-	44.2	-
11	35.4	2.09 (H11 ax), 2.30 (H11 eq)	33.3	2.02 (H11 ax), 2.27 (H11 eq)
12	79.6	4.09	78.4	4.33
13	84.3	-	85.0	-
14	141.1	5.90	147.2	6.05
15	112.5	4.99, 5.48	112.0	4.85, 4.90
16	21.6	1.14	19.6	1.20
17	108.8	4.78, 4.65	108.7	4.76, 4.67)
18	31.4	1.22	31.0	1.19
19	180.1	-	180.3	-
20	16.0	0.81	17.0	0.83

Table S2. Information on enzymes selected for study.

PDB	Enzyme	Class	PDB Binder	Resolution
5YHV	Aminotransferase	Transferase		2.70 Å
4YRN	Enoil-ACP reductase (InhA)	Oxidoreductase		2.55 Å

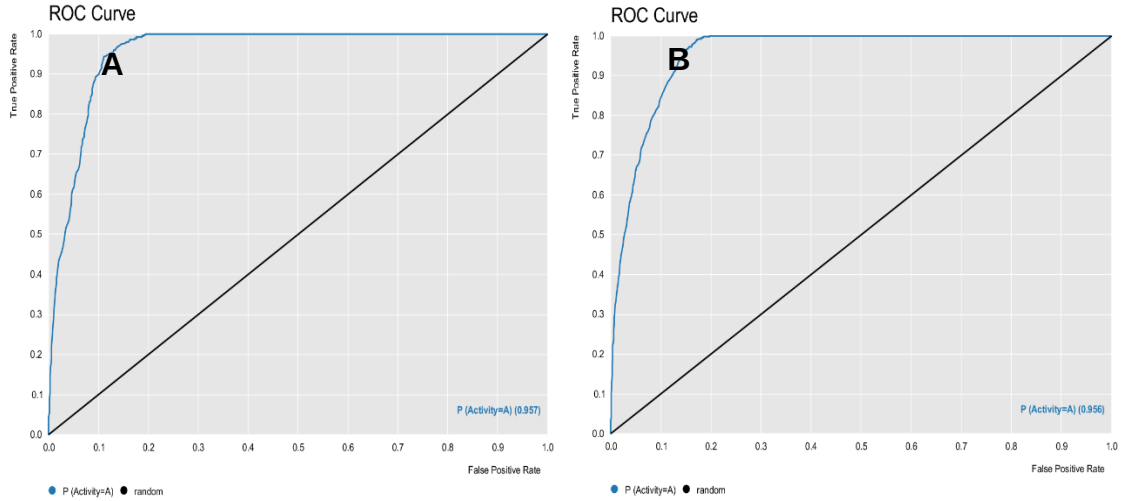


Figure S22. ROC curve generated for the RF model for *Mycobacterium tuberculosis*. (A) Test and (B) Cross-validation.

Table S3. Prediction of anti-tuberculosis biological activity for diterpenes.

ID	Probability	Activity	Domain of applicability
1	0.46	I	reliable
2	0.37	I	reliable
6	0.525	A	reliable

Table S4. Molecular docking results for the diterpenes and the crystallographic ligand in the two enzymes.

ID	Moldockscore	
	Aminostransferase	**InhA**
1	-74.142	-75.8373
2	-36.5368	-56.7605
6	-53.7718	-89.4165
Ligante PDB	-98.9217	-43.532

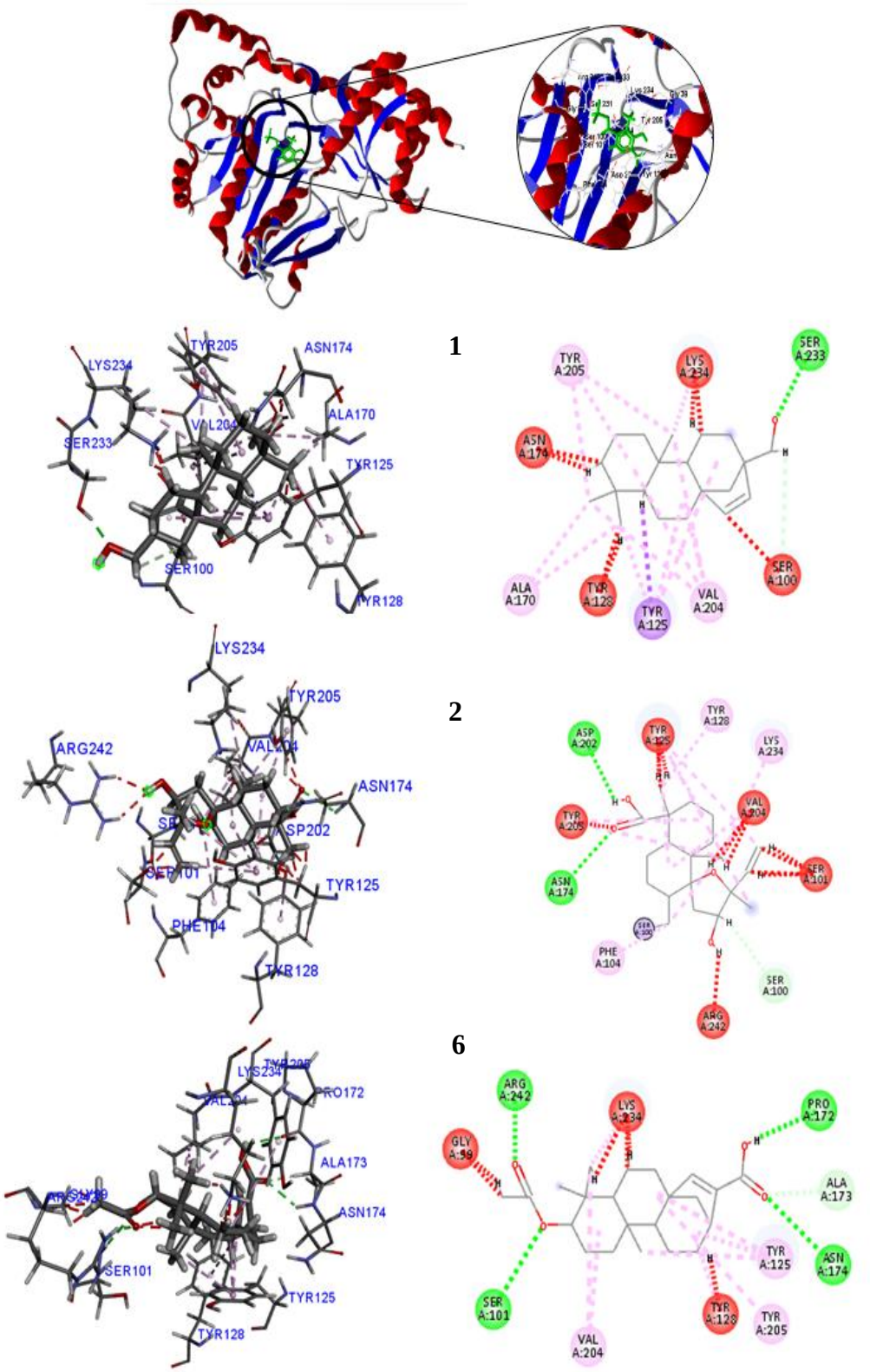


Figure S23. 3D and 2D interactions between diterpenes and the Aminotransferase enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink and electrostatic interactions are highlighted in red.

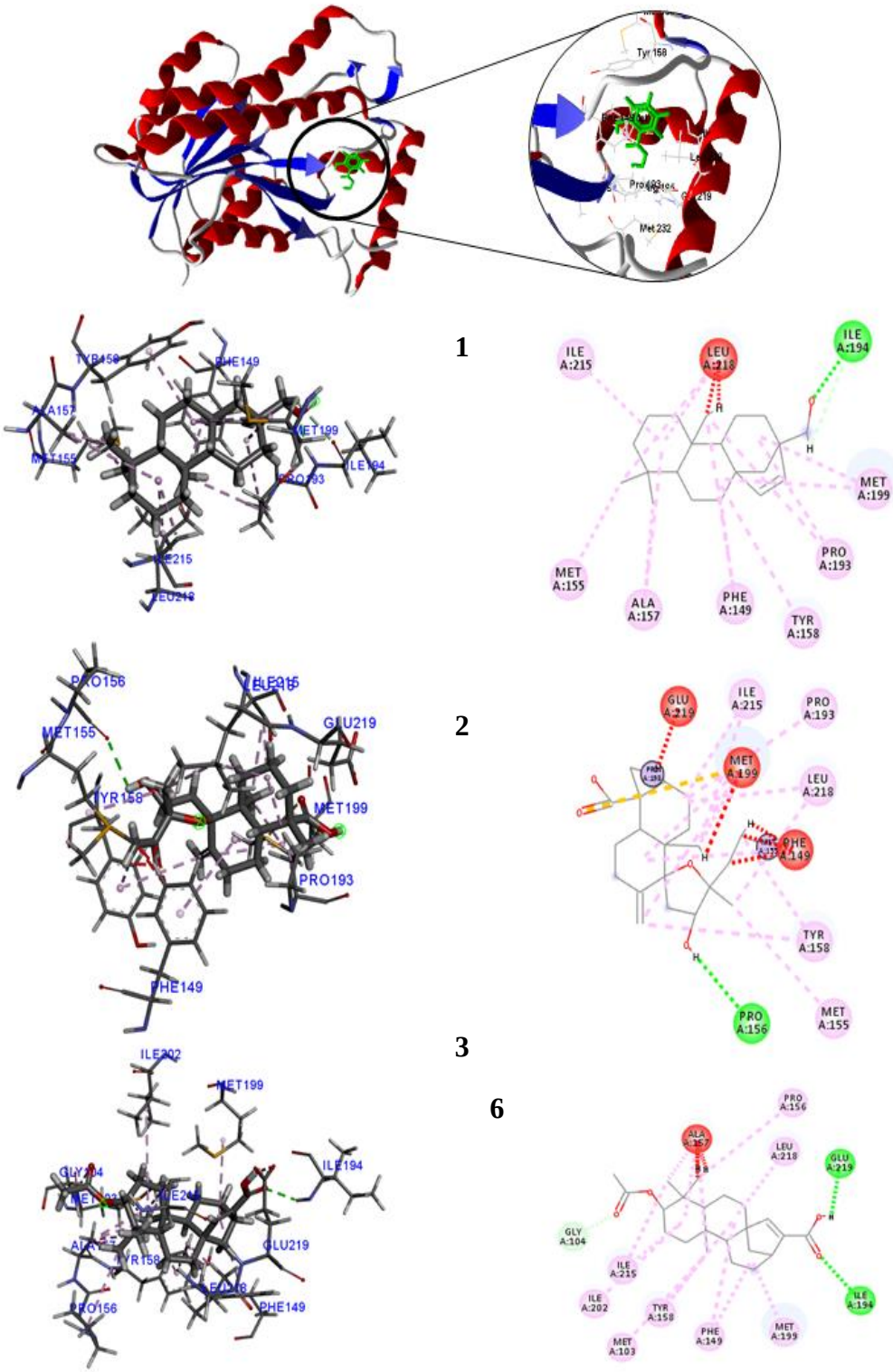


Figure S24. 3D and 2D interactions between the diterpenes and the InhA enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink and electrostatic interactions are highlighted in red.

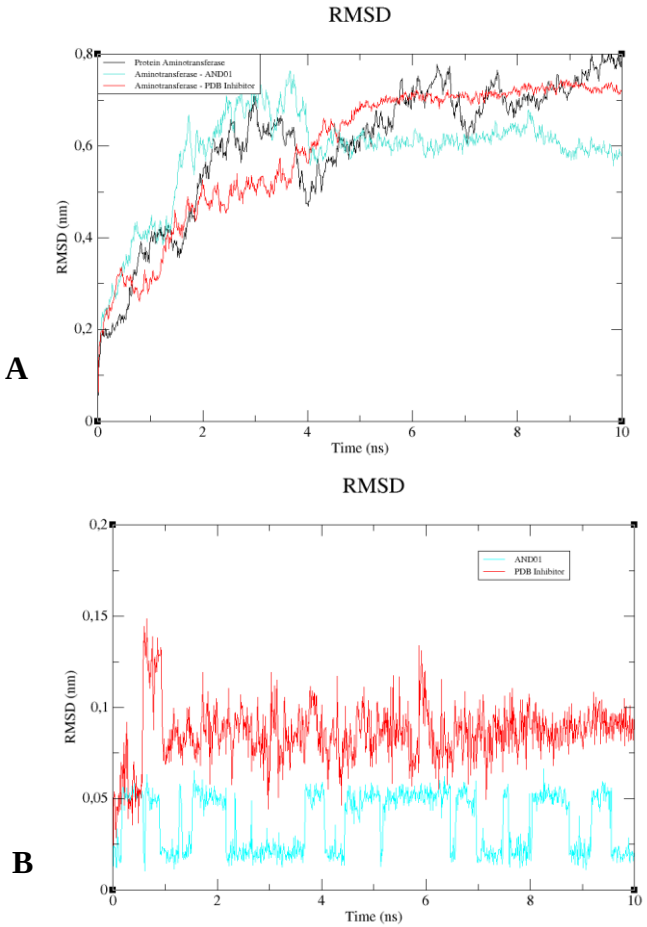
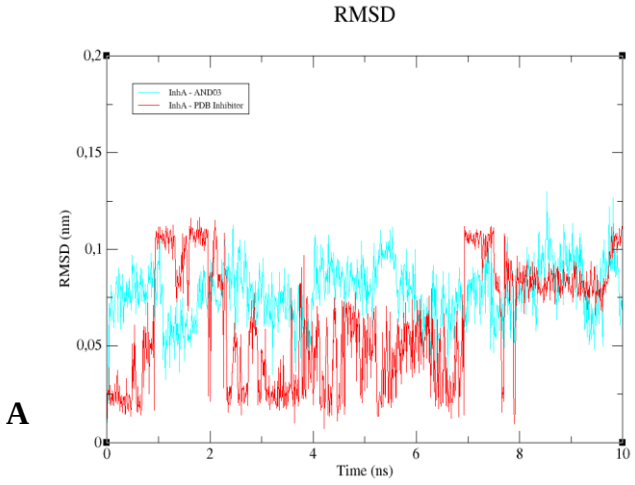


Figure S25. The RMSD values for the Cα atoms for the complexed Aminotransferase enzyme and for the diterpenes and the PDB ligand. (A) Aminotransferase complexed to diterpenes and PDB ligand. (B) **2** compound and the PDB linker.



B

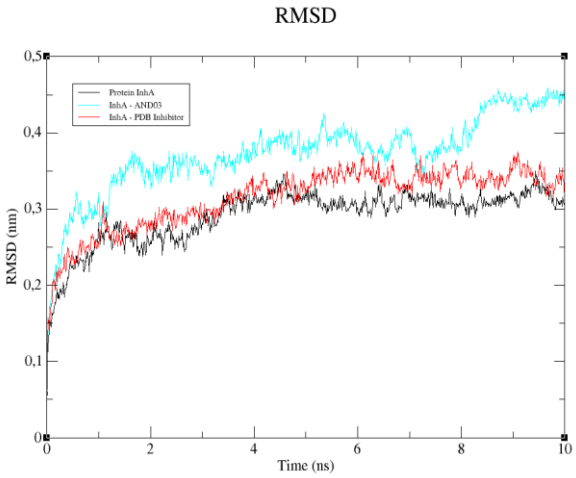


Figure S26. The RMSD values for the Cα atoms for the complexed InhA enzyme and for the diterpenes and the PDB ligand. (A) InhA complexed to diterpenes and PDB ligand. (B) 6 compound and the PDB linker.

Anexo III

Material suplementar do artigo: Four diterpenes identified *in silico* were isolated from Hyptidinae and demonstrated in vitro activity against *Mycobacterium tuberculosis*

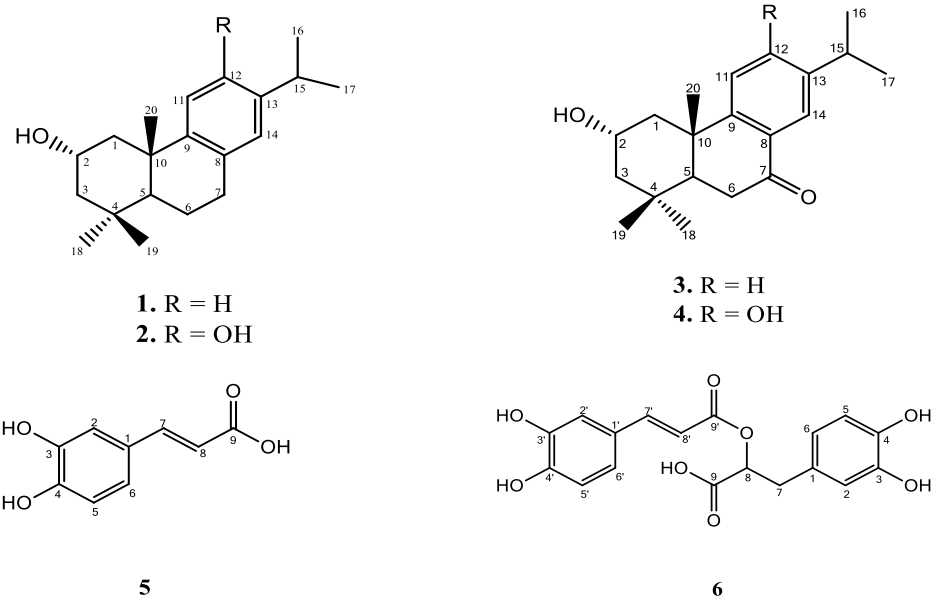


Figure S1. Chemical constituents isolated from the leaves of *Mesosphaerum sidifolium*. Pomiferin D (1), Salviol (2), Pomiferin E (3), 2α-Hydroxysugiol (4), Caffeic acid (5), Rosmarinic acid (6).

Compound **1** was obtained in the form of an amorphous white powder. In the ^{13}C NMR spectrum, 18 signals were detected, which correspond to 20 carbon atoms. Of these, five signals were attributed to nonhydrogenated carbons, and methyl carbons each, six signals were attributed to methinic carbons, and four signals were attributed to methylene carbons (Supplementary material, Figure S2). The correlation of CH_3-20 with C-9 assigned this carbon unequivocally, not allowing chemical shift inversion with C-14 as proposed in the literature (Ulubelen and Topcu 1992). Furthermore, the correlation of δ_H 1.19 ppm (CH_3-20) with δ_C 47.9 ppm unequivocally assigned this carbon to C-1, allowing an exchange of C-1 with C-3, correcting what was proposed in the literature (Supplementary material Table S1) (Ulubelen and Topcu 1992). This proposal is further reinforced by the correlation of the signal δ_H 0.98 ppm (CH_3-19) with the carbon at δ_C 50.9 ppm, assigned to C-3 (Supplementary material, Figure S3). The direct correlation of H-11 with δ_C 123.9 ppm attributes this carbon to C-11 and corrects what is proposed in the literature, which assigns this carbon to C-14 (Ulubelen and Topcu 1992) (Supplementary material, Figure S4). Finally, with C-1, C-3, C-6 and C-7 assigned, the hydroxyl was inserted on C-2. This proposal was reinforced by the

correlations of the signal at δ_H 2.62 ppm (H-1) with δ_C 65.6 ppm, and δ_H 1.82 ppm (H-3) with this same carbon. In the COSY experiment, correlations of the signals at δ_H 7.17 ppm (H-11) with δ_H 6.98 ppm (H-12), and δ_H 2.80 ppm (H-15) with δ_H 1.21 ppm (3H-16 and 3H-17) were observed, confirming the aromatic system and the isopropyl unit (Supplementary material, Figure S5). The correlation of δ_H 4.03 ppm (H-2) with δ_H 1.33 ppm (H-1) and δ_H 1.82 ppm (H-3) was also observed, confirming H-2 and reinforcing the insertion of the hydroxyl on this carbon. The relative stereochemistry of 1 was determined through the NOESY analysis. The hydroxyl inserted on C-2 was shown by the correlation of the signal at δ_H 4.03 ppm with the signal at δ_H 0.98 ppm (CH_3-19), establishing this hydroxyl as α-equatorial (Supplementary material, Figure S6 and S7).

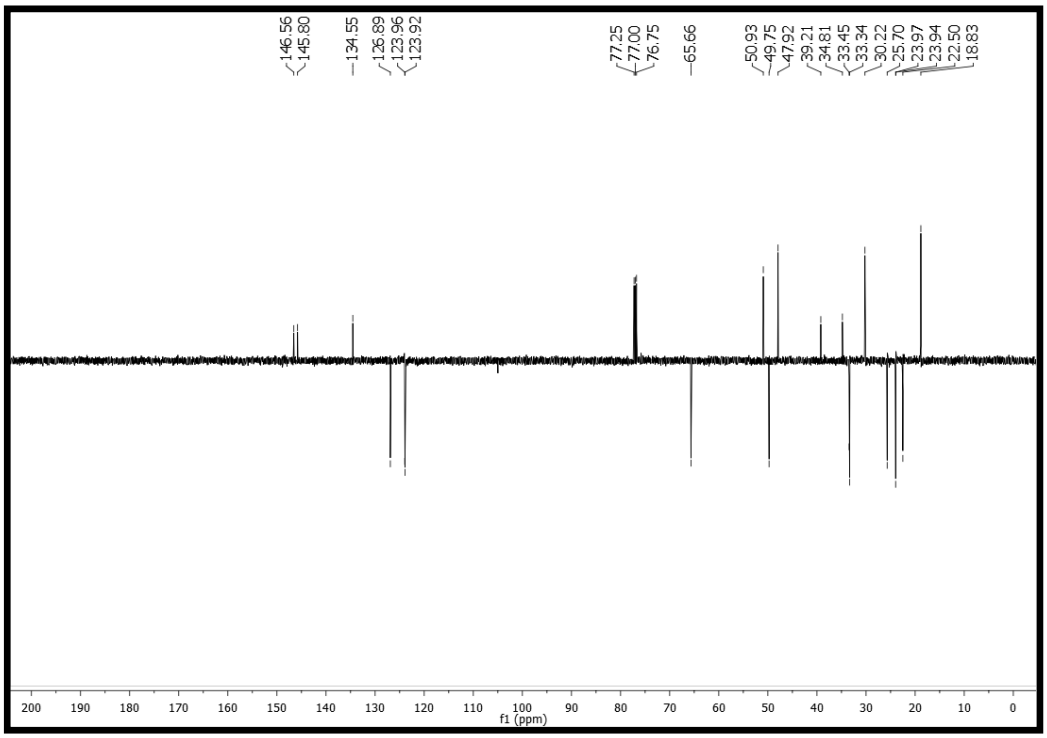


Figure S2. ^{13}C NMR APT spectrum (500 MHz, $CDCl_3$) of Pomiferin D (**1**).

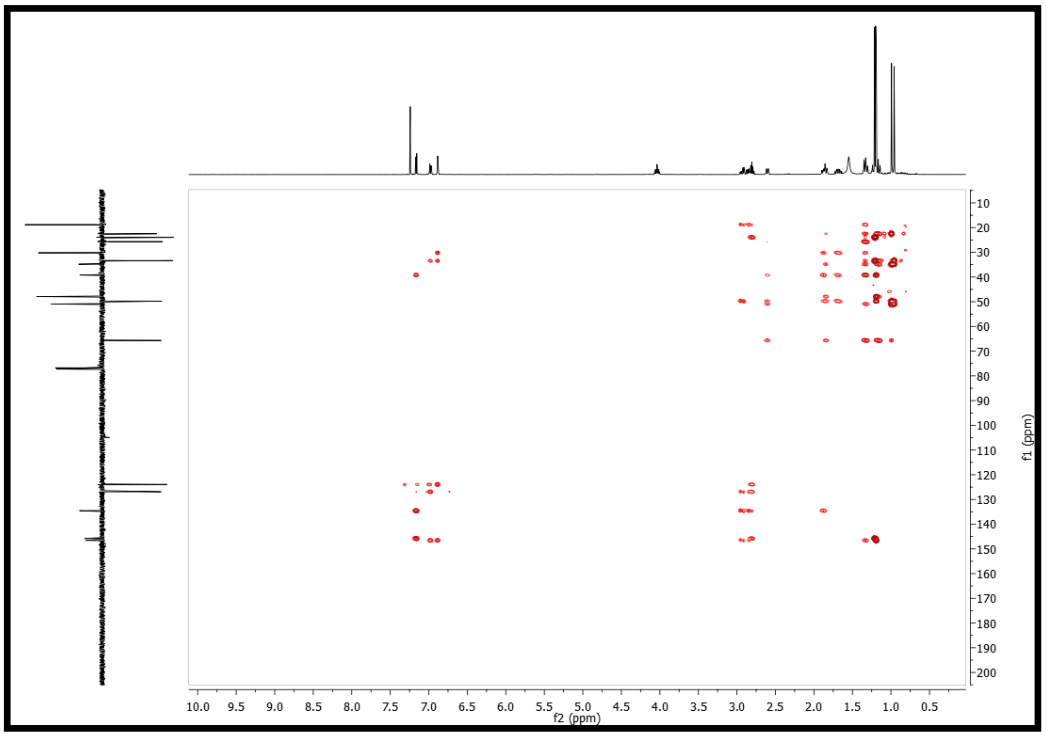


Figure S3. ^{1}H x ^{13}C spectrum HMBC (500 e 125 MHz, $CDCl_3$) of Pomiferin D (**1**).

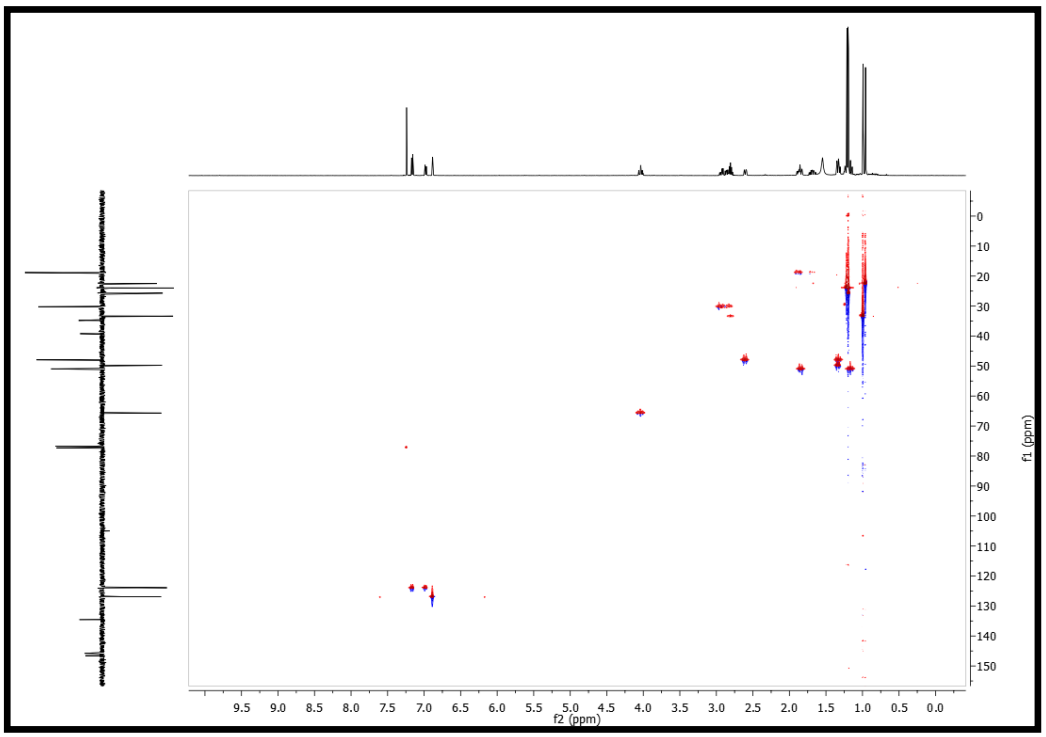


Figure S4. ^{13}C NMR HMQC correlation spectrum (500 e 125 MHz, $CDCl_3$) of Pomiferin D (**1**).

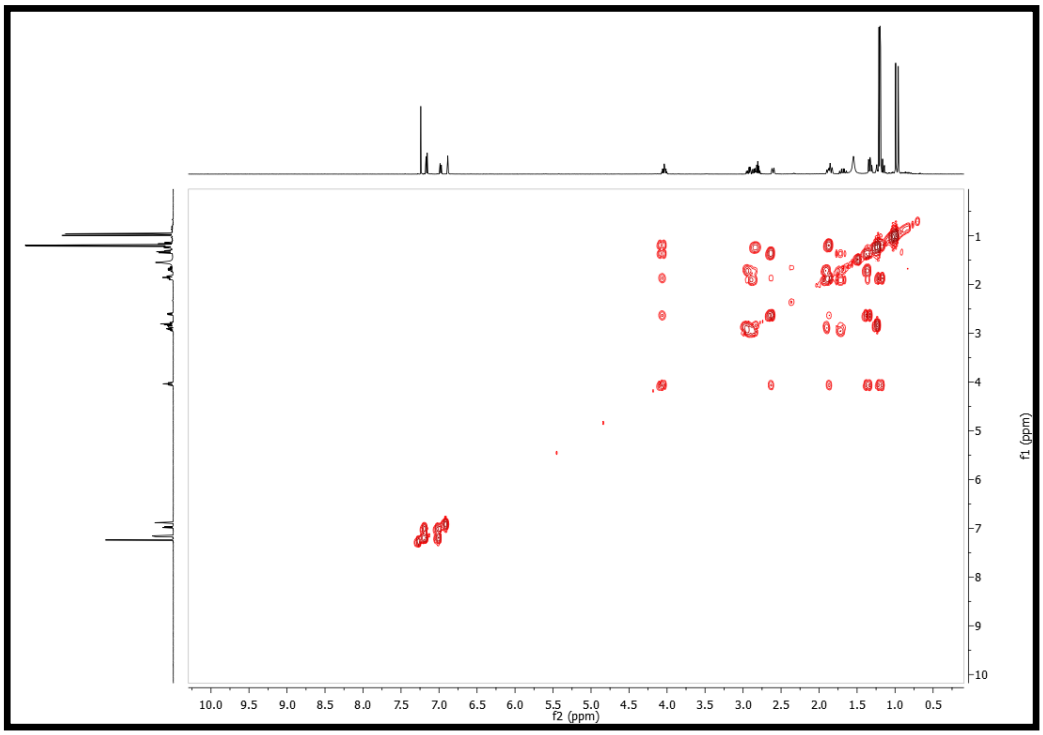


Figure S5. ^{1}H x ^{1}H COSY spectrum (500 MHz, $CDCl_3$) of Pomiferin D (**1**).

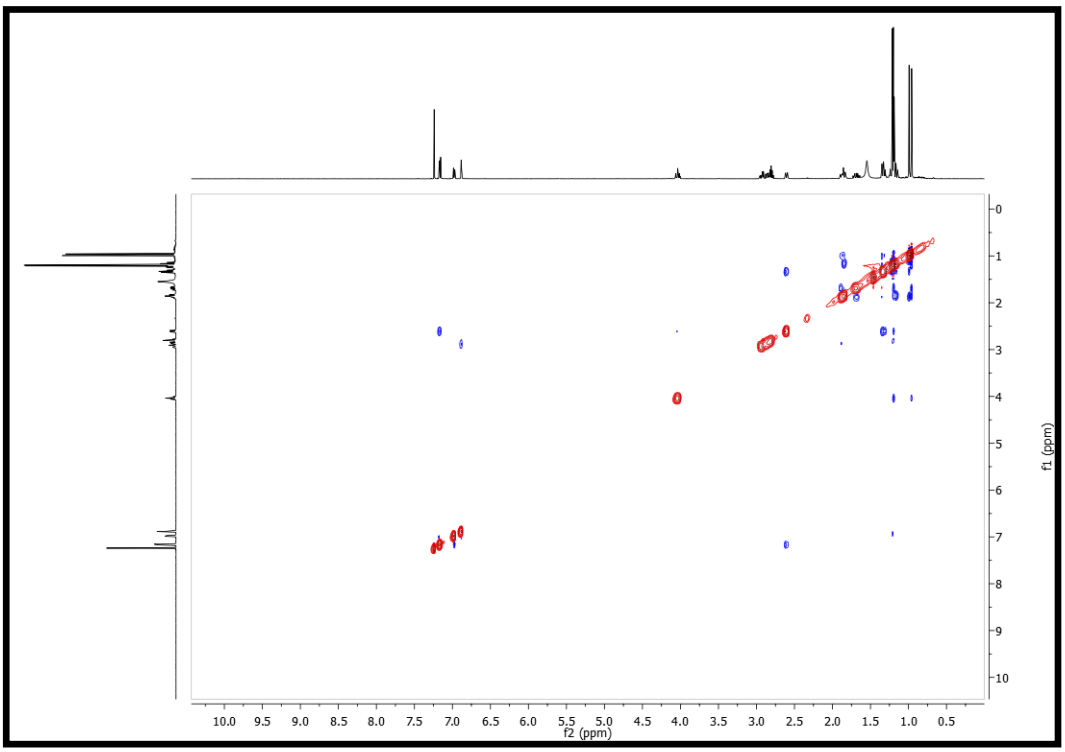


Figure S6. ^{1}H x ^{1}H NOESY spectrum (500 MHz, $CDCl_3$) of Pomiferin D (**1**).

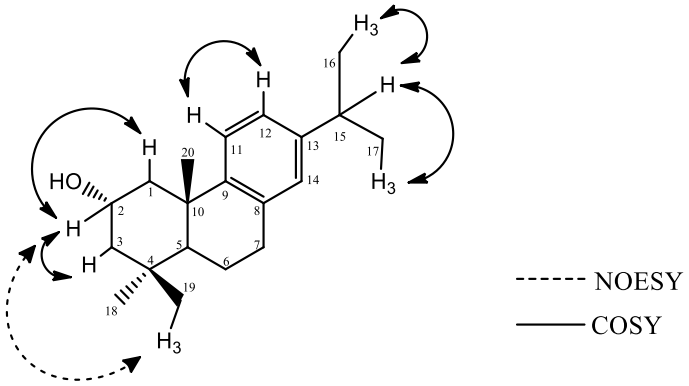


Figure S7. Stereochemistry established by NOESY and correlations of COSY of Pomiferin D (**1**).

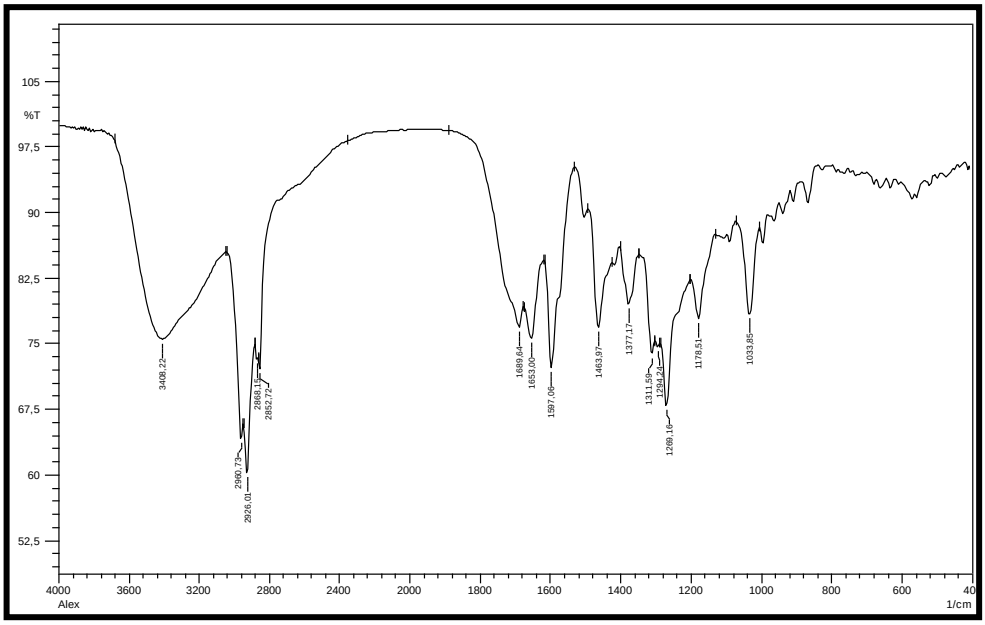


Figure S8. IR spectrum of 2α-hydroxysugiol (**4**) obtained in tablet of KBr.

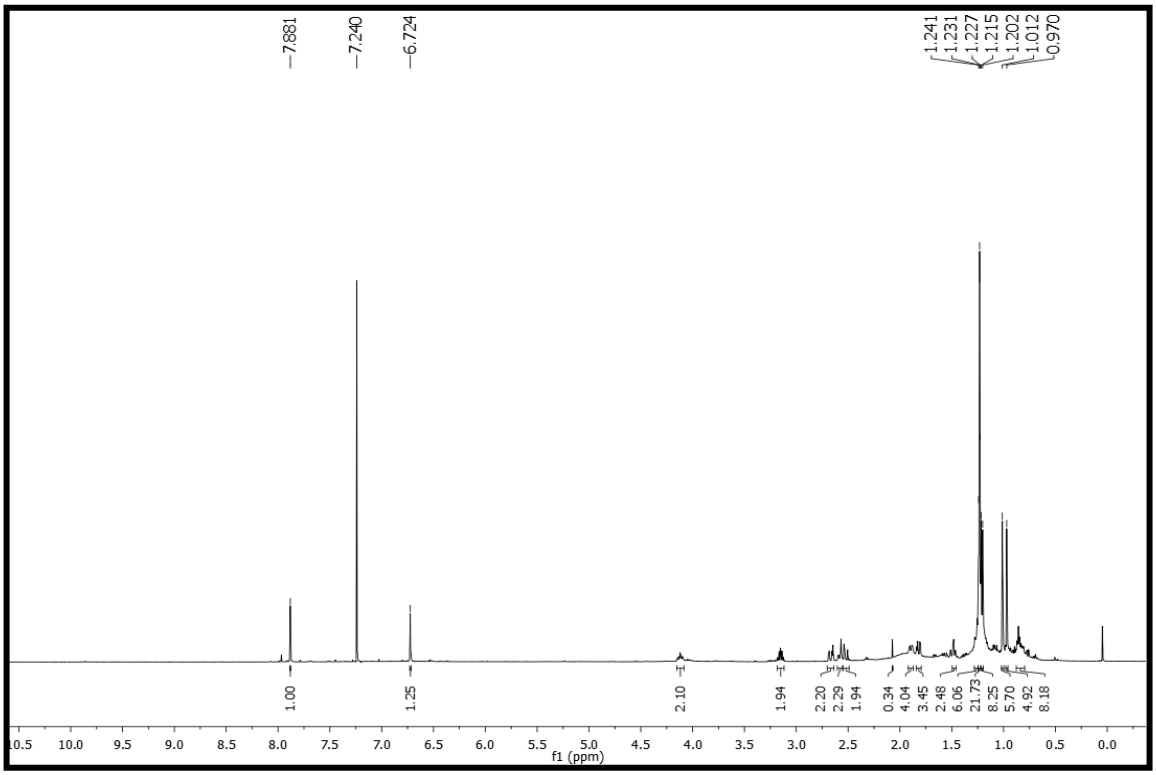


Figure S9. NMR ^{1}H spectrum (500 MHz, $CDCl_3$) 2α-hydroxysugiol (**4**).

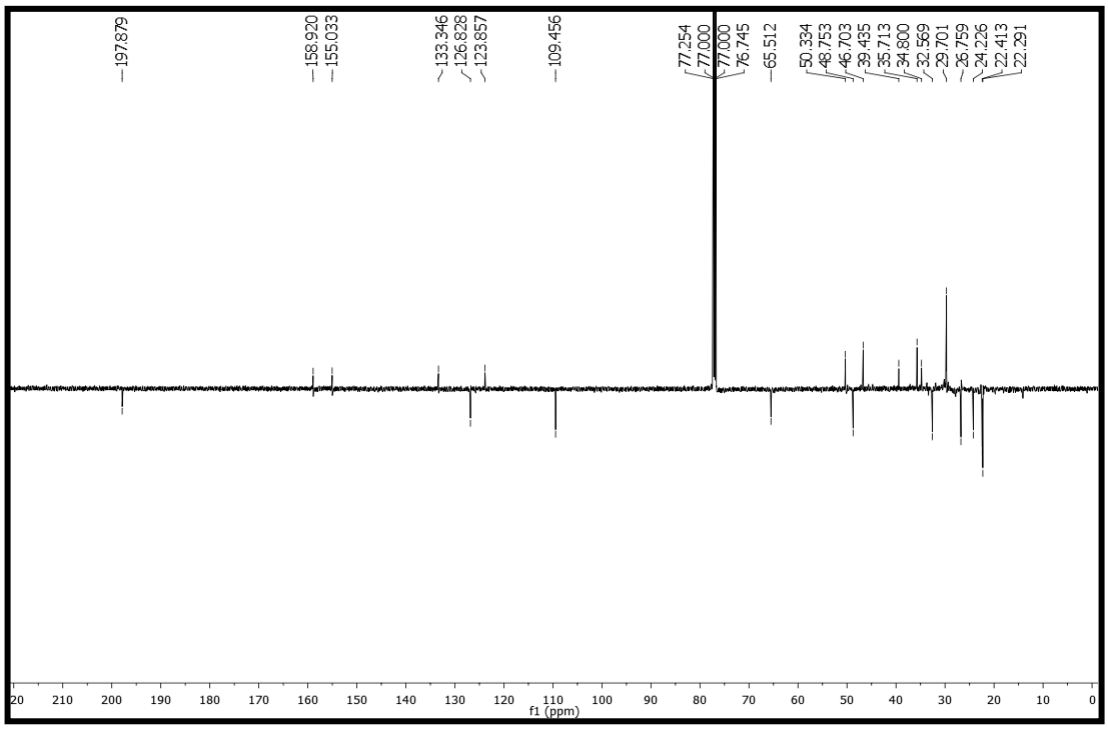


Figure S10. ^{13}C NMR APT spectrum (125 MHz, $CDCl_3$) of 2α-hydroxysugiol (**4**).

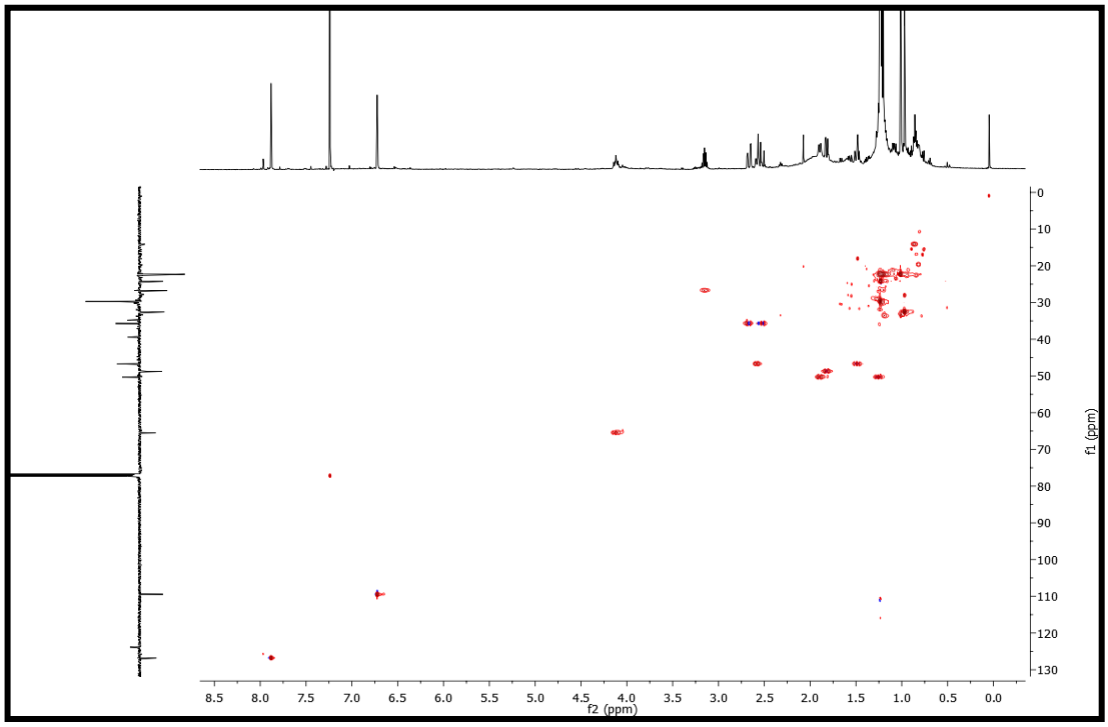


Figure S11. ^{1}H x ^{13}C spectrum HMQC (500 e 125 MHz, $CDCl_3$) of 2α-hydroxysugiol (**4**).

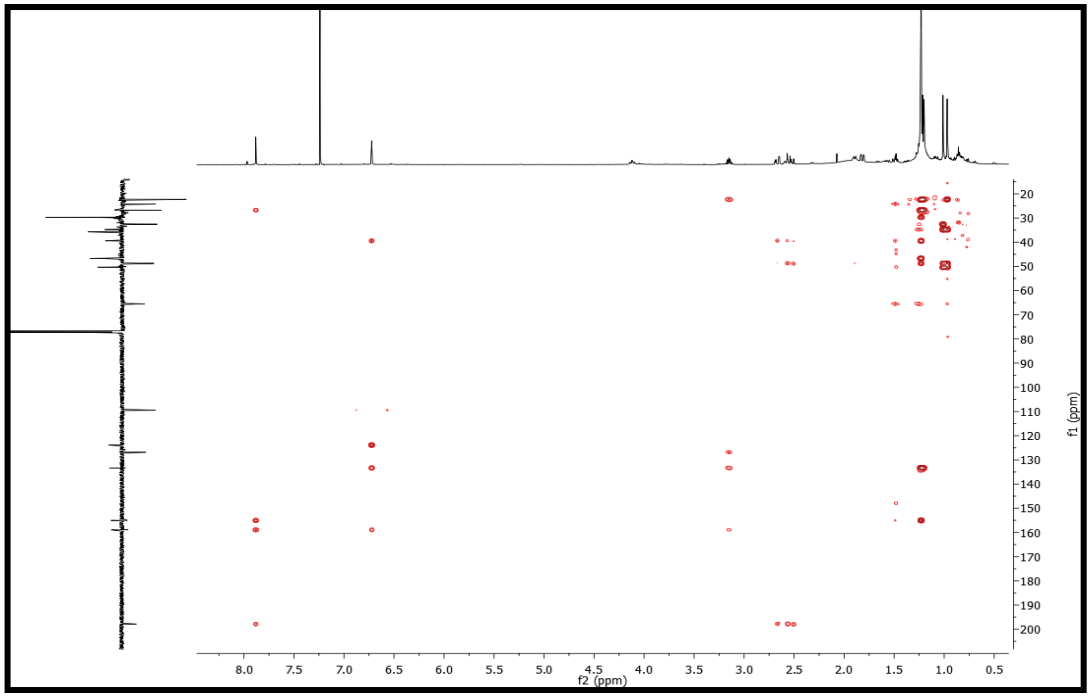


Figure S12. ^{1}H x ^{13}C spectrum HMBC (500 e 125 MHz, $CDCl_3$) of 2α-hydroxysugiol (**4**).

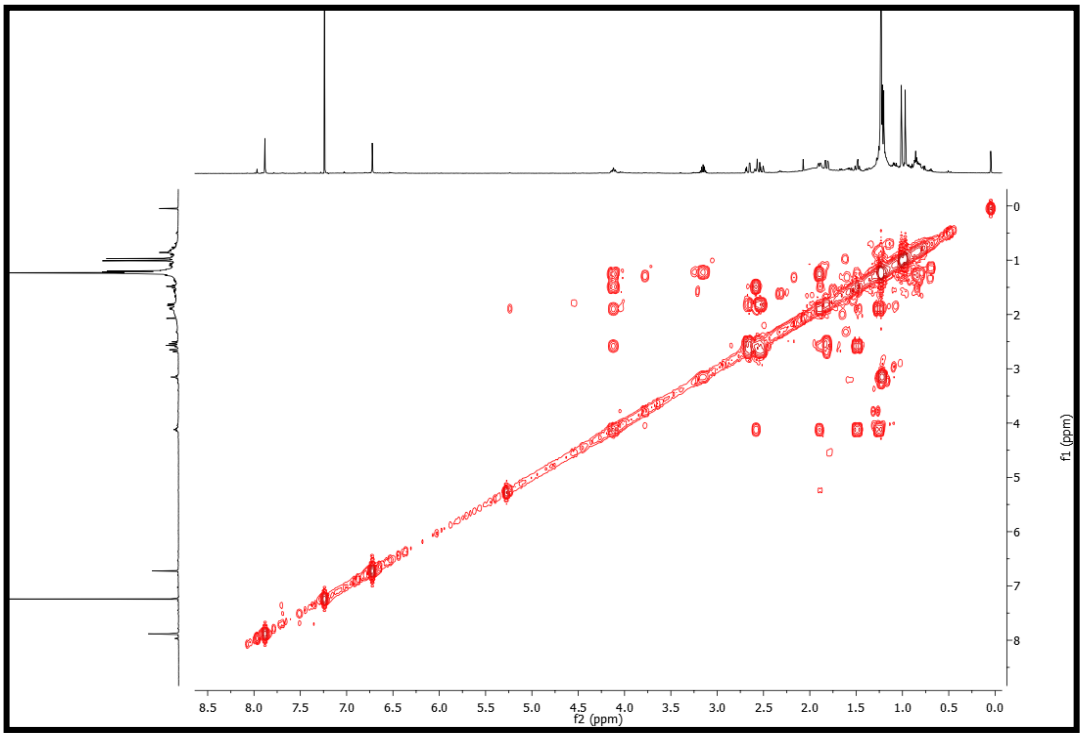


Figure S13. COSY spectrum (500 MHz, $CDCl_3$) of 2α-hydroxysugiol (**4**).

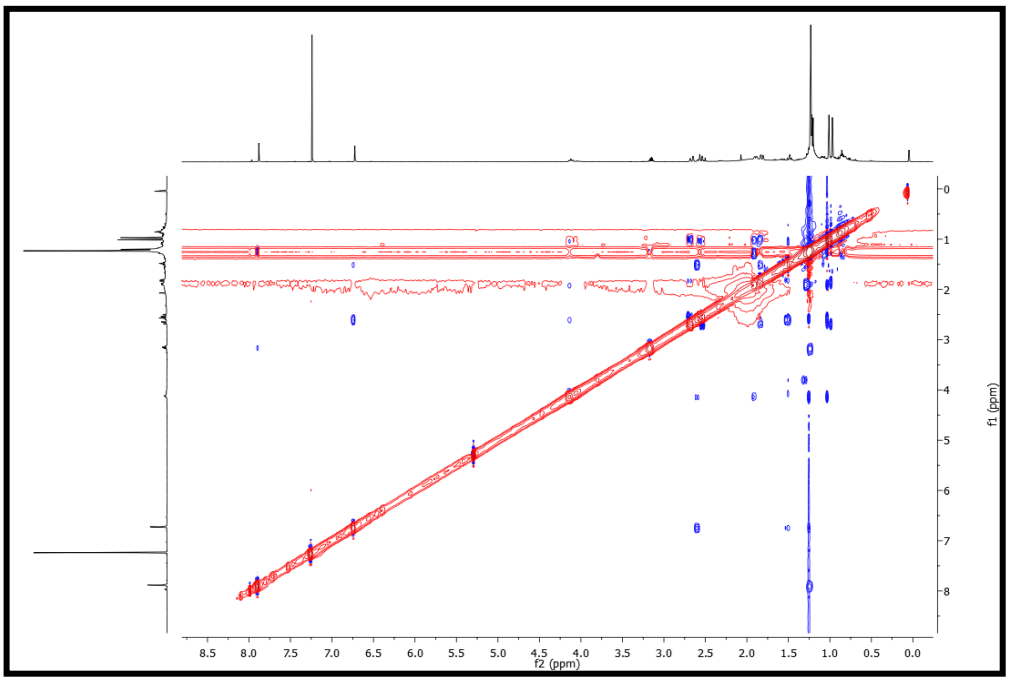


Figure S14. NOESY spectrum (500 MHz, $CDCl_3$) of 2α-hydroxysugiol (**4**).

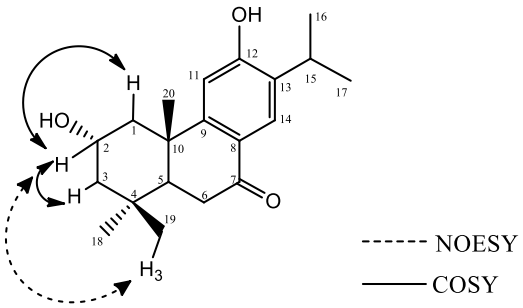


Figure S15. Stereochemistry established by NOESY and COSY correlations of 2α-hydroxysugiol (**4**).

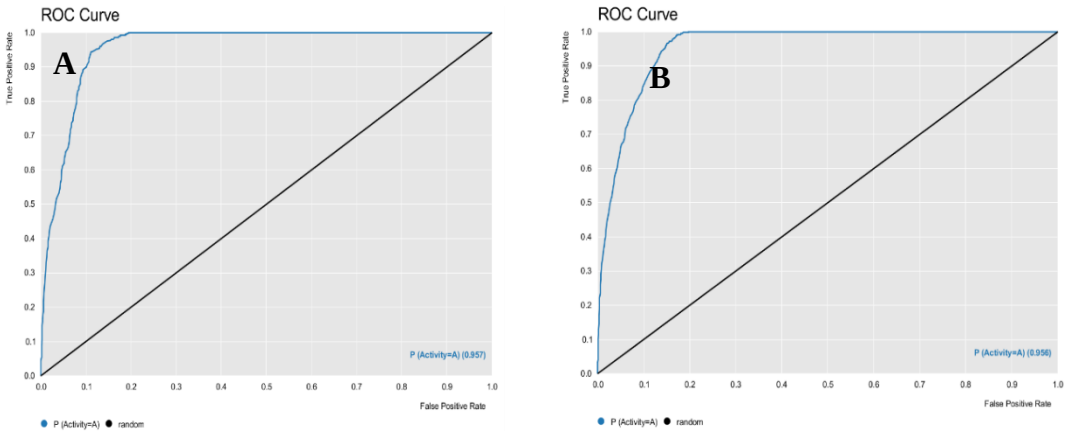


Figure S16. Graphical representation of the ROC curves of the internal cross-validation (A) and the test (B) group.

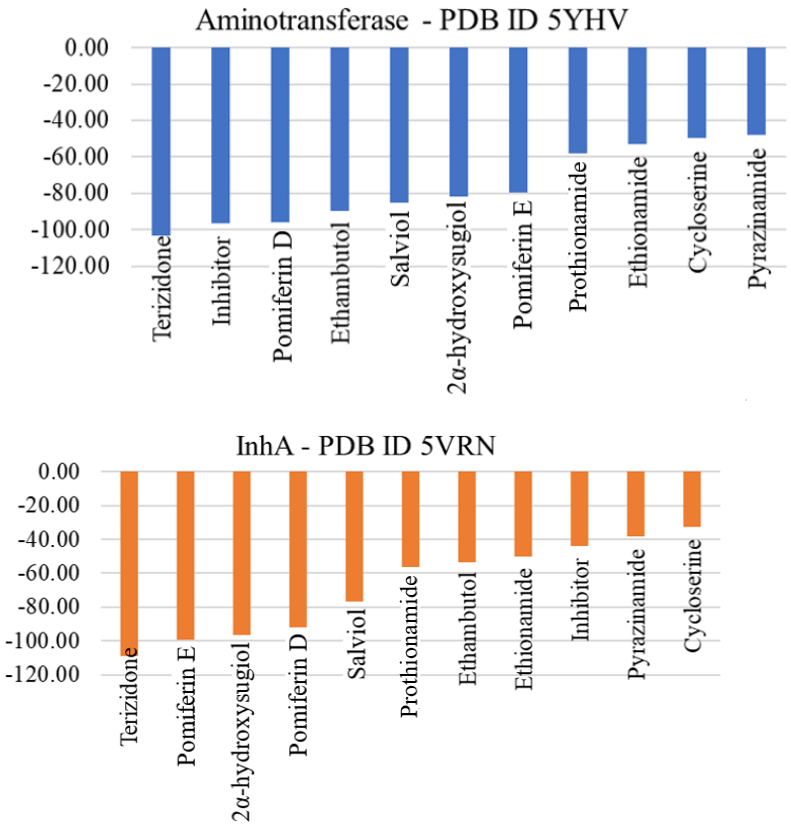
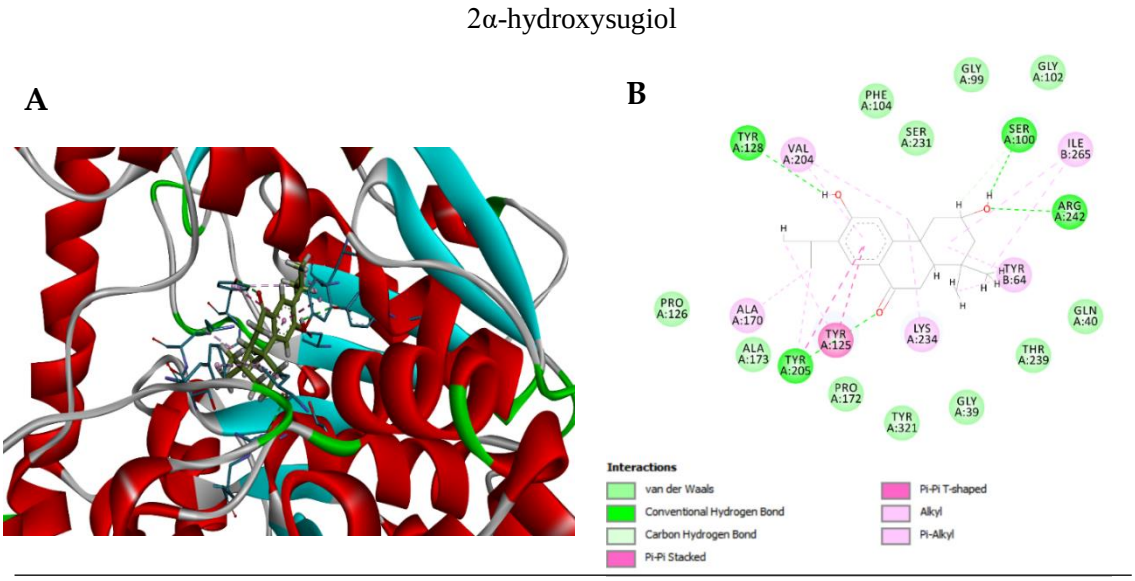
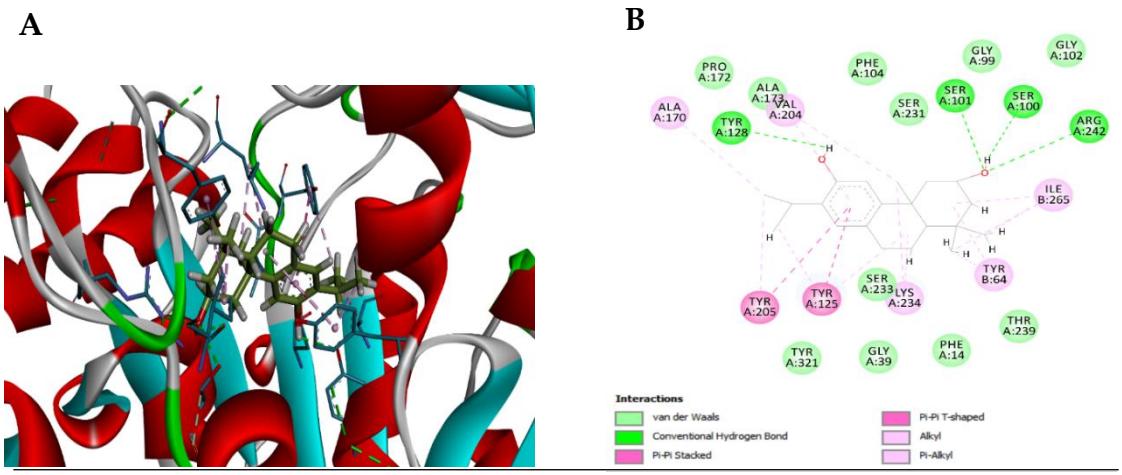
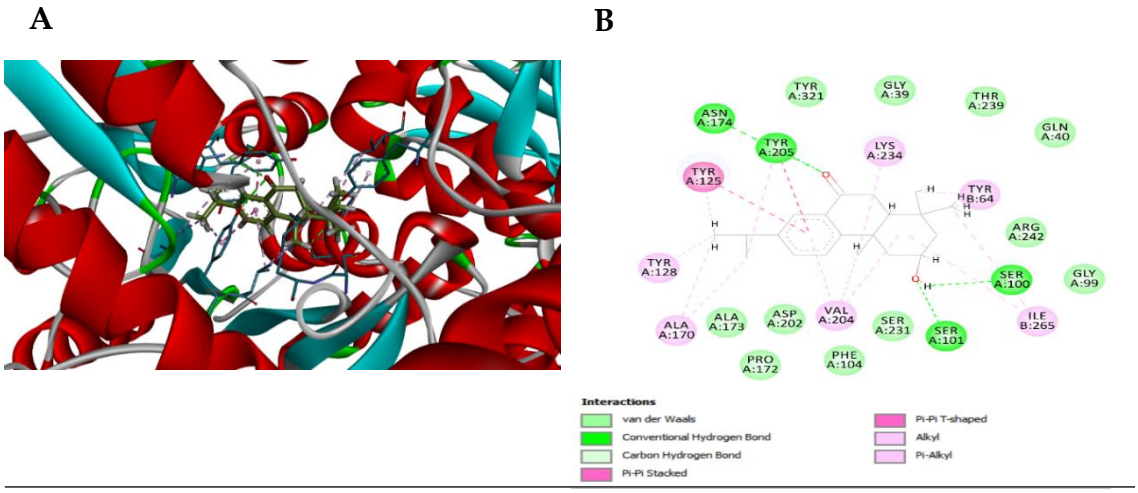


Figure S17. Ranking of the inhibitor energies complexed with the protein, four diterpenes analyzed and the six controls used, for each protein.





Pomiferin E



Pomiferin D

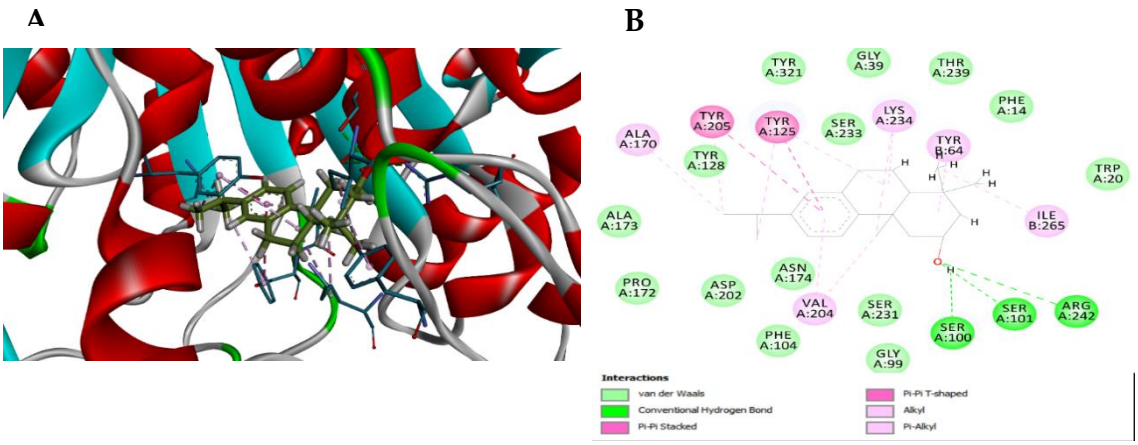
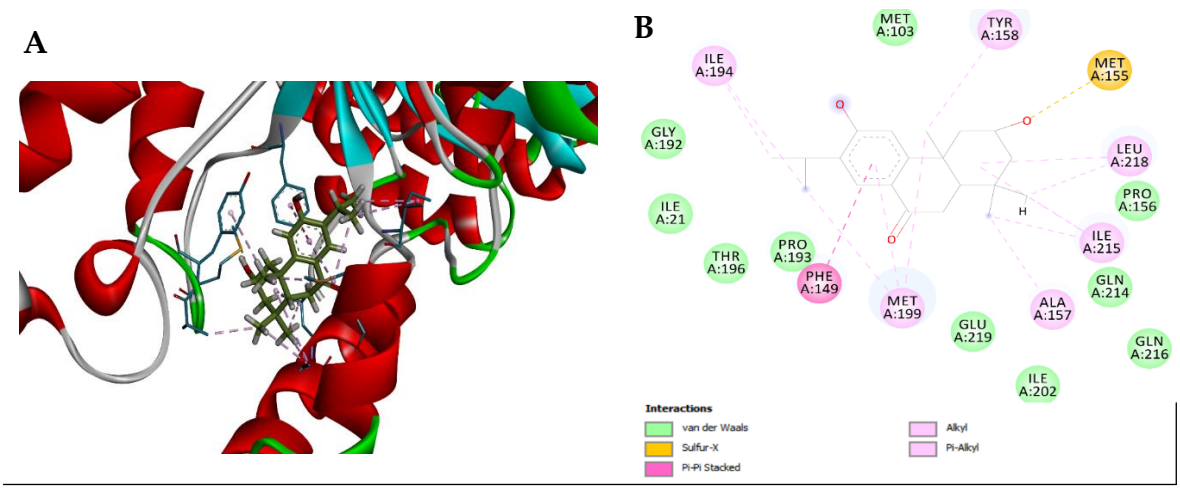
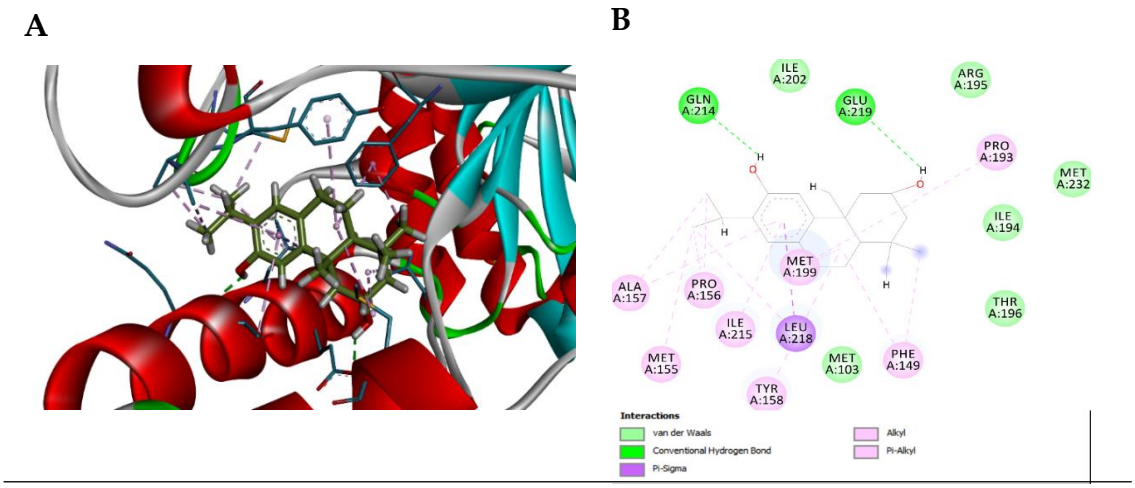


Figure S18. A) 3D interactions; B) 2D interactions with analyzed diterpenes and their interactions with aminotransferase (PDB ID 5YHV).

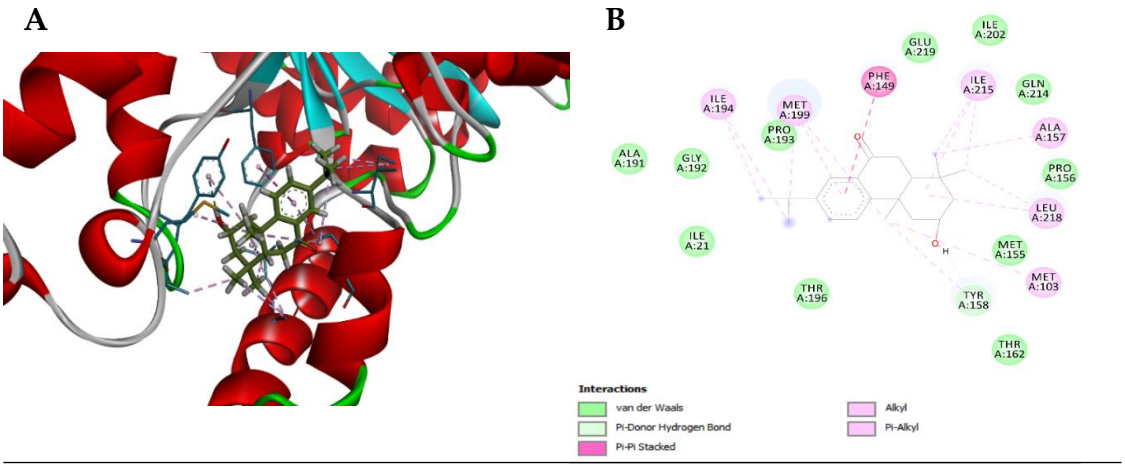
2α-hydroxysugiol



Salviol



Pomiferin E



Pomiferin D

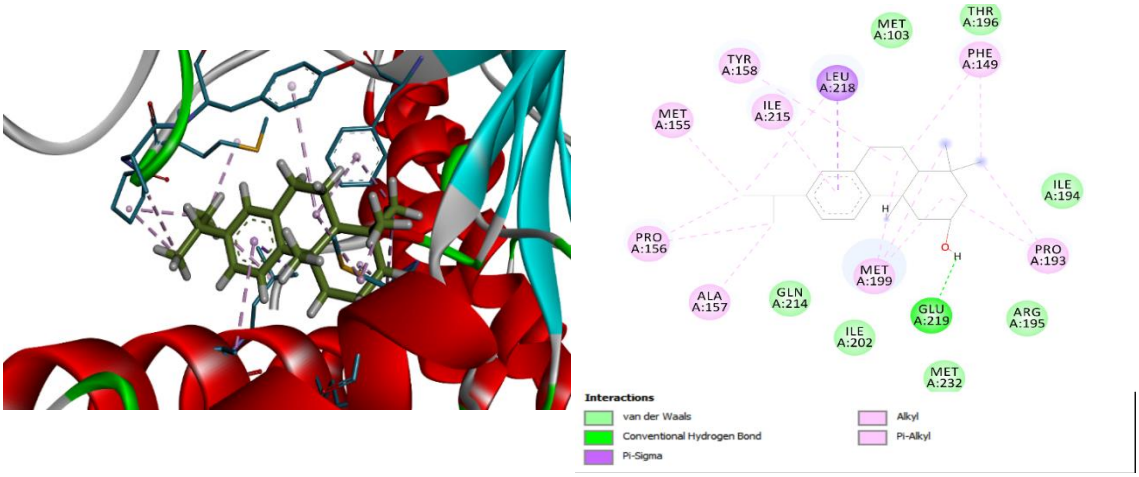
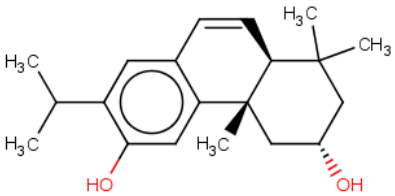
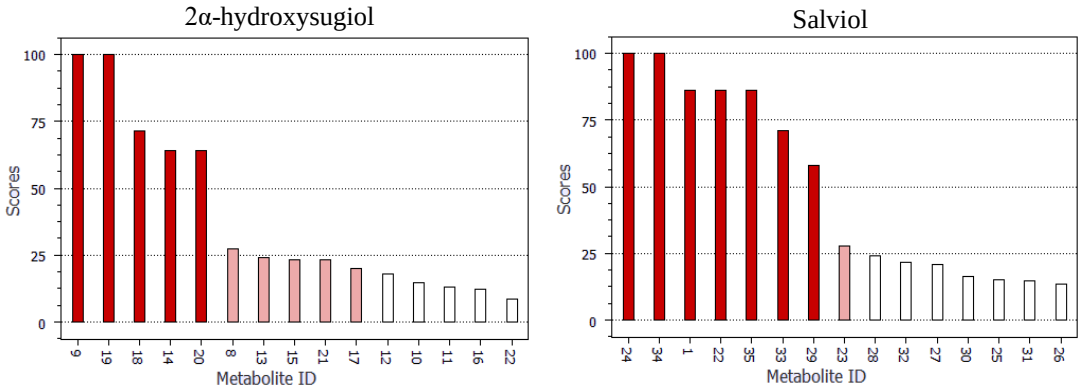


Figure S19. A) 3D interactions; B) 2D interactions with analyzed diterpenes and their interactions with InhA oxidoreductase (PDB ID 5VRN).



ID	Salviol
ID MET	MET35
Reaction	Dehydrogenation
Score	85.93
TOX	High – ESR

Figure S20. Hepatic metabolite with a risk of high toxic effect on the reproductive system.



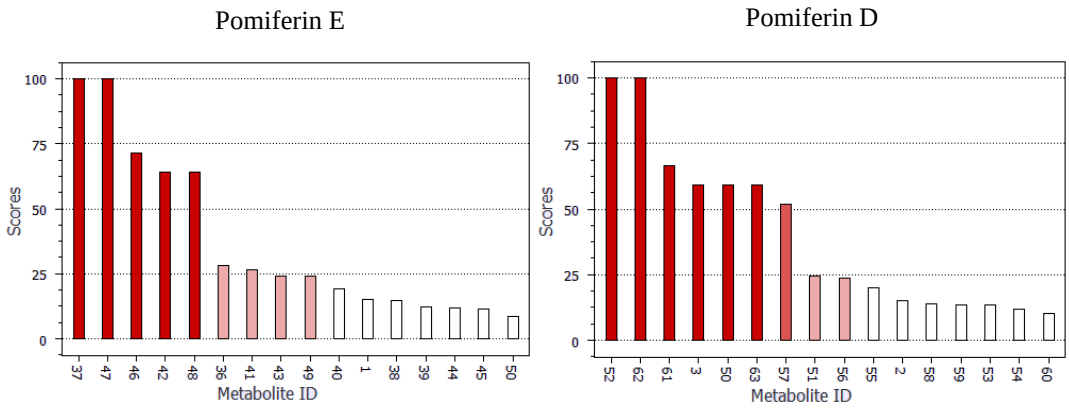


Figure S21. Hepatic metabolite with a risk of high toxic effect on the reproductive system.

Table S1. ^{1}H NMR data ($CDCl_3$, 500 MHz) and ^{13}C (125 MHz), heteronuclear results. Chemical shifts (δ ppm) and coupling constants (*J* Hz) of Pomiferin D (**1**).

C	$\delta^{1}H$	$\delta^{13}C$	HMBC (H→C)
1	1,33 eq (m) 2,62 (dt, *J* = 12,0; 3,0)	47,9	C2, C3, C5, C10
2	4,03 (tt, *J* = 15,0; 4,0)	65,6	
3	1,82 eq (dd, *J*=2,5; *J* = 4,0) 1,6 ax (m)	50,9	C2, C4, C19, C18, C5, C1
4	--	34,8	
5	1,33 (m)	49,7	C3, C4, C6, C7, C18, C19, C20
6	1,82 (m)	18,8	C4, C5, C7, C10
7	2,92 ax (m) 2,88 eq (dd, *J* = 8,0; 11,0)	30,2	C5, C6
8	--	134,5	
9	--	146,5	
10	--	39,2	
11	7,17 (d), *J* = 8,0)	123,9	C8, C10, C13
12	6,98 (dd, *J* = 8,0; 2,0)	123,9	C9, C11, C14, C15
13	--	145,8	
14	6,88 (sl)	126,9	C7, C9, C12, C15
15	2,80 (m)	33,4	C16, C17
16	1,21 (d, *J* = 7,0)	23,9	C13, C15, C17
17	1,21 (d, *J* = 1,0)	23,9	C13, C15, C17

18	0,95 (s)	22,5	C3, C4, C5, C19
19	0,98 (s)	33,3	C3, C4, C5, C18
20	1,19 (s)	25,6	C1, C4, C9, C10

[a] Assigned according to the correlations observed in Heteronuclear Multiple Bond Coherence (HMBC) and/or Heteronuclear Single Quantum Coherence (HSQC).

Table S2. ^{13}C NMR (125 MHz) data of Pomiferin D (**1**) and Salviol (**2**).

C	**1**	**2**
1	47,9	47,8
2	65,6	65,0
3	50,9	50,7
4	34,8	34,8
5	49,7	49,6
6	18,8	18,9
7	30,2	29,5
8	134,5	126,2
9	146,5	147,2
10	39,2	37,2
11	123,9	110,9
12	123,9	151,2
13	145,8	132,8
14	126,2	126,0
15	33,4	26,7
16	23,9	22,5
17	23,9	22,7
18	22,5	22,4
19	33,3	33,3
20	25,6	25,6

Table S3. ^{13}C NMR (125 MHz) data of Pomiferin D (**1**) and Pomiferin E (**3**).

C	**2**	**3**
1	47,9	46,8
2	65,6	65,0
3	50,9	50,5
4	34,8	34,7
5	49,7	48,6

6	18,8	35,9
7	30,2	199,3
8	134,5	130,3
9	146,5	152,6
10	39,2	39,4
11	123,9	123,4
12	123,9	125,1
13	145,8	147,0
14	126,2	132,6
15	33,4	33,6
16	23,9	23,8
17	23,9	23,7
18	22,5	22,2
19	33,3	32,5
20	25,6	24,4

Table S4. ^{1}H NMR ($CDCl_3$, 500 MHz) and ^{13}C (125 MHz) data, heteronuclear correlation results. Chemical shifts (δ ppm) and coupling constants (*J* Hz) of 2α-hydroxysugiol (**4**).

C	δ ^{1}H	δ^{13}C	HMBC (H→C)
1	2,59 (m, H-1a) 1,48 (d, *J* = 4,0 Hz, H-1b)	46,7	C-20, C-10, C-3, C-2, C-5
2	4,09 (m, H-2*β*)	65,5	--
3	1,87 (m, H-3a); 1,24 (m, H-3b)	50,3	C-4, C-2, C-18
4	--	34,8	--
5	1,83 (dd, *J* = 14,0; 3,5 Hz)	48,7	C-19, C-20, C-18, C-6, C-10
6	2,69 (dd, *J* = 4,0; 18,0 Hz, H-6a); 2,50 (s, H-6b)	35,7	C-4, C-10, C-5, C-7
7	--	197,8	--
8	--	123,8	--
9	--	155,0	--
10	--	39,4	--
11	6,72 (s)	109,4	C-10, C-8, C-13, C-12
12	--	158,9	-
13	--	133,3	-
14	7,88 (s)	126,8	C-15, C-9, C-12, C-7
15	3,18 (m)	26,7	C-12, C-13, C-14, C-17
16	1,09 (s)	22,4	C-17, C-15

17	1,24 (s)	22,2	C-16, C-15, C-13
18	0,97 (s)	22,2	C-19, C-4, C-5, C-3
19	1,01 (s)	32,5	C-18, C-4, C-5, C-3
20	1,22 (s)	24,2	C-10, C-1, C-5, C-9

Table S5. Statistical parameters of the antitubercular activity prediction model.

Parameter	Cross validation	Test
Sensitivity	0.96	0.97
Specificity	0.84	0.85
PPV	0.78	0.78
NPV	0.97	0.98
MCC	0.78	0.80
Accuracy	0.89	0.89

Table S6. Diterpenes of the subtribe Hyptidinae (Lamiaceae) in order of prediction of assets to inactive ones, domain of applicability (AD), activity probability and popular name. Presenting plant data, as part of the plant and geographic location, available at SistematX (https://sistematx.ufpb.br).

Rank	AD	Prob	ATV	Popular name	Species	Part of the plant	Geographic location (*in decimal*)
1	reliable	0.74	A	Hyptisolide A	*Hyptis crenata*	Aerial parts	-
2	reliable	0.73	A	*11,14-dihydroxy-12-methoxy-7-oxo-8,11,13-abietatrien-19,20-olide	*Gymneia platanifolia*	Roots	-7.220782, -39.411246
3	reliable	0.7	A	*19,20-epoxy-12-methoxy-11,14,19-trihydroxy-7-oxo-8,11,13-abietatriene	*Gymneia platanifolia*	Roots	-7.220782, -39.411246
4	reliable	0.695	A	Incanone	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
5	reliable	0.69	A	Rosmanol	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
6	reliable	0.68	A	7-O-methylrosmanol	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
7	reliable	0.67	A	Hyptol	*Eplingiella fruticosa*	Roots	-8.267183, -56.567744
8	reliable	0.655	A	Diacetylepimethylrosmanol	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
9	reliable	0.655	A	Esquirolin B	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
10	reliable	0.645	A	*7α,15,19-triacetoxy-2α-hydroxylabda-8(17),(13Z)-diene	*Cantinoa americana*	Aerial parts	19.182031, -96.177193

11	reliable	0.64	A	*11,12,16-trihydroxy-17(15→16)-abeo-abieta-8,11,13-trien-7-one	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
12	reliable	0.62	A	*6α,11,12,15-tetrahydroxy-8,11,13-abietatrien-7-one	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
13	reliable	0.615	A	Umbrosone	*Mesosphaerum sidifolium*	Roots	-8.510022, -35.738452
14	reliable	0.615	A	*15,19-diacetoxy-2α,7α-dihydroxylabda-8(17),(13Z)-diene	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
15	reliable	0.615	A	*19-acetoxy-2α,7α-dihydroxylabda-8(17),(13Z)-dien-15-al	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
16	reliable	0.615	A	*19-acetoxy-7α,15-dihydroxylabda-8(17),(13Z)-dien-2-one	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
17	reliable	0.615	A	19-oxo-inuroyleanol	*Gymneia platanifolia*	Roots	-7.220782, -39.411246
18	reliable	0.615	A	*3,7,11,15-tetramethylhexadec-1-en-3-ol	*Condea verticillata*	Aerial parts	
19	reliable	0.61	A	Coulterone	*Gymneia platanifolia*	Roots	-7.220782, -39.411246
20	reliable	0.61	A	Methoxynepetaefolin	*Condea undulata*	Aerial parts	-23.599130, -46.624186
21	reliable	0.61	A	*7-seco-7(20),11(20)-diepoxyabieta-8,11,13-triene	*Medusantha martiusii*	Roots	-7.369688, -40.083188

22	reliable	0.605	A	*19-acetoxy-2α,7α,15-trihydroxylabda-8(17),(13Z)-diene	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
23	reliable	0.6	A	(16S)-12,16-epoxy-11,14-dihydroxy-*17(15→16)-abeo-abieta-8,11,13-trien-7-one	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
24	reliable	0.6	A	Carnasol	*Medusantha martiusii*	Roots	-7.547858, -39.390218
25	reliable	0.6	A	14-methoxy-taxodione	*Eplingiella fruticosa*	Roots	-8.267183, -56.567744
26	reliable	0.595	A	Carnosol	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
27	reliable	0.595	A	15β-methoxyfasciculatin	*Condea undulata*	Aerial parts	-23.599130, -46.624186
28	reliable	0.585	A	Diacetylepiethylrosmanol	*Hyptis dilatata*	Aerial parts	8.088724, -80.978305
29	reliable	0.585	A	*19-acetoxy-2α,7α-dihydroxylabda-14,15-dinorlabd-8(17)-en-13-one	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
30	reliable	0.575	A	7-Acetyl-12-methoxyhorminone	*Condea verticillata*	Roots	18.324778, -77.030339
31	reliable	0.57	A	*11,12,15-trihydroxy-8,11,13-abietatrien-7-one	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
32	reliable	0.57	A	(-)-Salzol	*Hypenia salzmannii*	Leaves	-8.211531, -35.569429
33	reliable	0.57	A	12-methoxycarnosic acid	*Medusantha martiusii*	Roots	-7.369688, -40.083188
34	reliable	0.56	A	2α-hydroxysugiol	*Mesosphaerum sidifolium*	Aerial parts	-7.249827, -37.387660
35	reliable	0.555	A	15α-methoxyfasciculatin	*Condea undulata*	Aerial parts	-23.599130, -46.624186

36	reliable	0.545	A	Pomiferin D	*Mesosphaerum sidifolium*	Aerial parts	-7.249827, -37.387660
37	reliable	0.54	A	Salviol	*Mesosphaerum sidifolium*	Aerial parts	-7.249827, -37.387660
38	unreliable	0.54	A	*8,11,13-abietatriene	*Mesosphaerum suaveolens*	-	29.322153, 79.443199
39	reliable	0.53	A	Martiusane	*Medusantha martiusii*	Roots	-7.369688, -40.083188
40	reliable	0.53	A	*Abieta-7,13-dien-3-one	*Mesosphaerum suaveolens*	-	-21.424154, -45.935383
41	reliable	0.525	A	Pomiferin E	*Mesosphaerum sidifolium*	Aerial parts	-7.249827, -37.387660
42	reliable	0.525	A	Dehydroabietol	*Mesosphaerum suaveolens*	-	29.322153, 79.443199
43	reliable	0.515	A	**ent*-3β-acetoxy-kaur-15-eno-17-óico	*Leptohyptis macrostachys*	Aerial parts	-7.249827, -37.387660
44	reliable	0.515	A	*7,11,15-trimethyl-3-methylenehexadecane-1,2-diol	*Condea verticillata*	Aerial parts	
45	reliable	0.51	A	Inuroyleanol	*Gymneia platanifolia*	Roots	-7.220782, -39.411246
46	unreliable	0.51	A	Abietadiene	*Mesosphaerum suaveolens*	-	22.744834, 79.090180
47	reliable	0.5	A	Methyl suaveolate	*Eremostachys glabra*	Rhizomes	38.122020, 46.276989
48	unreliable	0.5	A	Atisirene	*Mesosphaerum suaveolens*	-	5.588123, 12.989821
49	reliable	0.49	I	Horminone	*Eplingiella fruticosa*	Roots	-8.267183, -56.567744
50	reliable	0.48	I	Rimuene	*Mesosphaerum suaveolens*	-	5.588123, 12.989821

51	reliable	0.475	I	*11,14-dihidroxy-8,11,13-abietatrien-7-one	*Medusantha martiusii*	Roots	-7.134132, -39.593884
52	reliable	0.47	I	*2α,7α,15,19-tetrahydroxy-*ent*-labda-8(17),(13Z)-diene	*Cantinoa americana*	Aerial parts	19.182031, -96.177193
53	reliable	0.465	I	Abietol	*Mesosphaerum suaveolens*	-	5.588123, 12.989821
54	unreliable	0.46	I	Biformene	*Mesosphaerum suaveolens*	-	5.588123, 12.989821
55	reliable	0.455	I	14-O-methylsuaveolic acid	*Mesosphaerum suaveolens*	-	15.023678, 102.036348
56	reliable	0.45	I	Suaveolol	*Eremostachys glabra*	Rhizomes	38.122020, 46.276989
57	unreliable	0.45	I	Phyllocladene	*Mesosphaerum suaveolens*	-	5.588123, 12.989821
58	reliable	0.445	I	*3,7,11,15-tetramethylhexadec-2-en-1-ol	*Condea verticillata*	Aerial parts	
59	reliable	0.43	I	11-oxomanoyloxide	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
60	reliable	0.43	I	Suaveolic acid	*Mesosphaerum suaveolens*	Aerial parts	24.037945, 90.399486
61	reliable	0.43	I	Isosuaveolic acid	*Mesosphaerum suaveolens*	Whole plant	15.023678, 102.036348
62	reliable	0.43	I	Phytol	*Mesosphaerum suaveolens*	-	5.588123, 12.989821
63	reliable	0.425	I	11β-hydroxymanoyloxide	*Oocephalus crassifolius*	Roots	-13.012701, -41.359255
64	reliable	0.41	I	8α,9α-epoxysuaveolic acid	*Mesosphaerum suaveolens*	Whole plant	15.023678, 102.036348
65	reliable	0.405	I	Eritroxilol B	*Leptohyptis macrostachys*	Aerial parts	-7.249827, -37.387660

66	reliable	0.405	I	Manoyl oxide	*Mesosphaerum suaveolens*	-	-21.424154, -45.935383
67	reliable	0.385	I	*9α,13α-epi-dioxiabiet-8(14)-en-18-ol	*Mesosphaerum suaveolens*	Leaves	9.656955, 8.087476
68	reliable	0.385	I	Hyptenol	*Leptohyptis macrostachys*	Aerial parts	-7.249827, -37.387660

* Diterpene presented by the chemical name, as it has no popular name

Table S7. Results of prediction of antitubercular activity.

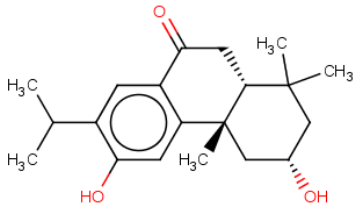
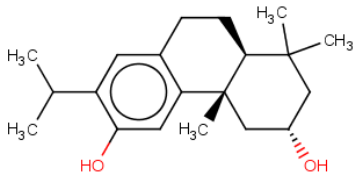
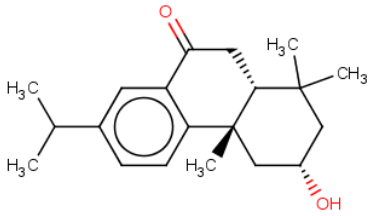
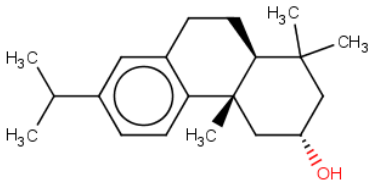
ID	Domain	ATV	%ATV	Structures
2α-hydroxysugiol	reliable	A	72	
Salviol	reliable	A	66.5	
Pomiferin E	reliable	A	67.5	
Pomiferin D	reliable	A	57.5	

Table S8. Percentage data of oral restriction, violation of the Lipinski rule and cytotoxicity.

ID	%ABS	LIP	TPSA	MUT	CAR	ESR	IRR	TOX
2α-hydroxysugiol	89.15	0	57.53	none	none	none	none	No
Salviol	95.04	1	40.46	none	none	none	none	No
Pomiferin E	96.13	0	37.30	none	none	none	none	No
Pomiferin D	102.02	1	20.23	none	none	none	none	No

LIP: Violations of Lipinski's rule;
TPSA: Total polar surface area;
%ABS: Oral absorption rate;
MUT: Mutagenicity;
CAR: Carcinogenicity;
ESR: Toxic effect on the reproductive system;
IRR: Skin irritability;
TOX: Total toxicity.

Table S9. Summary of the types of interaction and the amino acid residues of the aminotransferase involved.

Protein	ID	Energy [kj.mol^{-1}]	Interactions	
			Type	Waste
5YHV	Terizidone	-103.32	H-Bond	2-Ser100, Tyr125, 2-Tyr128, Ser231, 2-Ser233 and Arg242
			Eletrostatic	None
			Steric	4-Ala170, Pro172, 3-Tyr128, 2-Tyr125, Asp202, Ser100, 4-Asn174, 2-Lys234, 2-Arg243 and 2-Ser202
	Inhibitor	-96.70	H-Bond	Ser100, Ser231, Ser233, Ser101, 2-Arg242, Tyr205 and Asn174
			Eletrostatic	2-Arg242
			Steric	3-Lys234, Ser100, Arg242 and Ser101
	Pomiferin D	-96.08	H-Bond	Ser100, Arg242 and Ser101
			Eletrostatic	None
			Steric	Val204, Ser100, Tyr125, Asn174, 5-Lys234, 2-Tyr128 and 2-Asp202
	Ethambutol	-89.79	H-Bond	Ser100, Ser101, Lys234 and Tyr205
			Eletrostatic	None
			Steric	Ser100, Arg242, Lys234, Tyr174 and Tyr128
	Salviol	-85.09	H-Bond	Ser100, Ser101, Arg241, Tyr125 and Asp202
			Eletrostatic	None
			Steric	Val204, Ser101, Ser100, Arg242, Asn174, Ser233, 6-Lys234, Tyr125, 2-Asp202 and Tyr128
	2α-hydroxysugiol	-81.93	H-Bond	Ser100, Arg242, Ser101, Tyr205, Asn174, Tyr128 and Asp202
			Eletrostatic	None
			Steric	Ser233, Ser101, 2-Ser100, 3-Tyr125, 6-Lys234, Asn174, 2-Tyr128, 2-Asp202 and Val204
	Pomiferin E	-79.41	H-Bond	Ser100, Ser101, Tyr205 and Asn174

		Eletrostatic	None
		Steric	Asp202, Val204, Ser233, Ser101, 3-Ser100, Tyr125, Asn174, 6-Lys234 and 3-Tyr128

Table S10. Prediction data for possible liver metabolites.

ID	ID MET	Reaction	SCORE	MUT	CAR	ESR	IRR	TOX
2α-hydroxysugiol	MET9	Aliphatic Hydroxylation	100.00	none	none	none	none	No
	MET19	Dehydrogenation	100.00	none	none	none	none	No
	MET18	Alcoholic Oxidation	71.43	none	none	none	none	No
	MET14	Aliphatic Hydroxylation	64.29	none	none	none	none	No
	MET20	Dehydrogenation	64.29	none	none	none	none	No
	MET8	Aliphatic Hydroxylation	27.55	none	none	none	none	No
	MET13	Aliphatic Hydroxylation	23.98	none	none	none	none	No
	MET15	Aliphatic Hydroxylation	23.27	none	none	none	none	No
	MET21	Aliphatic Carbonylation	23.27	none	none	none	none	No
	MET17	Aromatic Hydroxylation	20.00	none	none	none	none	No
	MET12	Aliphatic Hydroxylation	17.86	none	none	none	none	No
	MET10	Aliphatic Hydroxylation	14.88	none	none	none	none	No
	MET11	Aliphatic Hydroxylation	12.94	none	none	none	none	No
	MET16	Aromatic Hydroxylation	12.08	none	none	none	none	No
	MET22	Ketone reduction	8.56	none	none	none	none	No
Salviol	MET24	Aliphatic Hydroxylation	100.00	none	none	none	none	No
	MET34	Dehydrogenation	100.00	none	none	none	none	No
	MET22	Aliphatic Hydroxylation	85.93	none	none	none	none	No
	MET35	Dehydrogenation	85.93	none	none	high	none	Yes
	MET33	Alcoholic Oxidation	71.11	none	none	none	none	No
	MET29	Aliphatic Hydroxylation	57.78	none	none	none	none	No
	MET23	Aliphatic Hydroxylation	27.68	none	none	none	none	No
	MET28	Aliphatic Hydroxylation	24.13	none	none	none	none	No
	MET32	Aromatic Hydroxylation	21.82	none	none	none	none	No
	MET27	Aliphatic Hydroxylation	20.74	none	none	none	none	No

	MET30	Aliphatic Hydroxylation	16.26	none	none	none	none	No
	MET25	Aliphatic Hydroxylation	15.25	none	none	none	none	No
	MET31	Aromatic Hydroxylation	14.80	none	none	none	none	No
	MET26	Aliphatic Hydroxylation	13.43	none	none	none	none	No
Pomiferim E	MET37	Aliphatic Hydroxylation	100.00	none	none	none	none	No
	MET47	Dehydrogenation	100.00	none	none	none	none	No
	MET46	Alcoholic Oxidation	71.43	none	none	none	none	No
	MET42	Aliphatic Hydroxylation	64.29	none	none	none	none	No
	MET48	Dehydrogenation	64.29	none	none	none	none	No
	MET36	Aliphatic Hydroxylation	28.04	none	none	none	none	No
	MET41	Aliphatic Hydroxylation	26.43	none	none	none	none	No
	MET43	Aliphatic Hydroxylation	24.00	none	none	none	none	No
	MET49	Aliphatic Carbonylation	24.00	none	none	none	none	No
	MET40	Aliphatic Hydroxylation	19.18	none	none	none	none	No
	MET38	Aliphatic Hydroxylation	14.88	none	none	none	none	No
	MET39	Aliphatic Hydroxylation	12.42	none	none	none	none	No
	MET44	Aromatic Hydroxylation	11.75	none	none	none	none	No
	MET45	Aromatic Hydroxylation	11.56	none	none	none	none	No
	MET50	Ketone reduction	8.60	none	none	none	none	No
Pomiferim D	MET52	Aliphatic Hydroxylation	100.00	none	none	none	none	No
	MET62	Dehydrogenation	100.00	none	none	none	none	No
	MET61	Alcoholic Oxidation	66.67	none	none	none	none	No
	MET50	Aliphatic Hydroxylation	59.26	none	none	none	none	No
	MET63	Dehydrogenation	59.26	none	none	none	none	No
	MET57	Aliphatic Hydroxylation	51.85	none	none	none	none	No
	MET51	Aliphatic Hydroxylation	24.34	none	none	none	none	No
	MET56	Aliphatic Hydroxylation	23.70	none	none	none	none	No
	MET55	Aliphatic Hydroxylation	19.89	none	none	none	none	No
	MET58	Aliphatic Hydroxylation	13.91	none	none	none	none	No
	MET59	Aromatic Hydroxylation	13.58	none	none	none	none	No

	MET53	Aliphatic Hydroxylation	13.46	none	none	none	none	No
	MET54	Aliphatic Hydroxylation	11.95	none	none	none	none	No
	MET60	Aromatic Hydroxylation	10.26	none	none	none	none	No

MUT: mutagenicity; CAR: carcinogenicity; ESR: toxic effect on the reproductive system.

IRR: irritability on the skin and TOX: total toxicity compared to the four parameters analyzed.

Table S11. Prediction data for ADME properties

ID	GI absorption[1]	BBB Permeant	P-gp substrate[2]	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	Log k_p (cm/s)[3]
2α-hydroxysugiol	High	Yes	Yes	No	No	No	Yes	No	-5.33
Salviol	High	Yes	No	No	No	No	Yes	No	-4.47
Pomiferin E	High	Yes	No	No	No	No	Yes	No	-4.97
Pomiferin D	High	Yes	No	No	No	No	Yes	No	-4.11

[1] Gastrointestinal absorption
[2] P-glycoprotein substrate
[3] Skin permeation

3. Experimental

3.1. Instrumentation and general procedures

Medium pressure liquid chromatography (MPLC) was performed on a BÜCHI Pump Manager C-615/605. Column chromatography (CC) and MPLC techniques were carried out using silica gel ART 7734 from Merck (Darmstadt, Germany) with particle size between 0.063 and 0.200 mm as stationary phases. Thin-layer chromatography (TLC) was performed on Whatman Flexible Aluminum-Backed TLC Plates AL SIL G/UV obtained from Merck (Darmstadt, Germany). The compounds applied in the chromatoplates were revealed by exposure to ultraviolet light from a UVGL-58 Mineralight device with two wavelengths (254 and 366 nm) and by impregnating the plates in glass vats saturated with iodine vapor. Organic solvents (hexane, dichloromethane, ethyl acetate and methanol) were used for the chromatography. Solvents from Merck (Darmstadt, Germany) were used for elution and deuterated solvents were used for Nuclear Magnetic Resonance (NMR). Infrared (IR) spectra were obtained using BOMEM MB-100 (Quebec, Canada) spectrometers in the range 400–4000 cm^{-1}, using KBr tablets (0.5 mg of sample/15 mg of KBr). The NMR spectra were

recorded in VARIAN SYSTEM spectrometers (California, EUA), operating at 500 MHz for proton (^{1}H NMR) and 125 MHz for carbon-13 (^{13}C NMR).

3.2. Plant material

The botanical material (aerial parts) of *M. sidifolium* (L'Hérit.) Harley & J.F.B. Pastore was collected in the semiarid region of Matureia - PB, Brazil, in August 2009, and identified by Prof. Maria de Fátima Agra, Ph.D. An exsiccate is kept with in Prof. Lauro Pires Xavier Herbarium, at the UFPB Campus of João Pessoa, identified as AGRA et al. 6964.

3.3. Isolation of chemical constituents

The plant material (2 kg of aerial parts) was dehydrated in forced air ovens operated at 40 °C for 72 h. After this, the material was pulverized in a mechanical mill, resulting in 1 kg of plant powder. It was then macerated with EtOH (95%) for 72 h in a percolator and this process was performed three times producing an ethanolic solution. The extractive solution was concentrated under reduced pressure at 40 °C, resulting in 100 g of crude ethanolic extract (CEE). Part of the CEE (10 g) was submitted to MPLC, using a silica gel stationary phase and pure or binary mixtures of hexane, ethyl acetate and methanol, in increasing degree of polarity, as the mobile phase. This resulted in 87 fractions, analyzed using TLC and combined according to their chromatographic profiles. These fractions were submitted to new chromatographic fractionation in silica gel, using binary mixtures of hexane, ethyl acetate and methanol in increasing degree of polarity as eluents, leading to the isolation of six compounds: Pomiferin D (**1**, 10 mg), Salviol (**2**, 20 mg), Pomiferin E (**3**, 84.5 mg), 2α-hydroxysugiol (**4**, 86.6 mg), caffeic acid (**5**, 32.8 mg) and rosmarinic acid (**6**, 95.1 mg). The 14–15 fraction (220 mg) was submitted to a CC, resulting in 98 fractions, which were analyzed using TLC and brought together according to their respective Rf values, in the ratio of hexane: acetate (90:10). Compound **1** represented fraction 9 obtained from this CC. Compound **2** was specifically obtained from the direct fractionation of CEE (fraction 17). Fraction 16 (300 mg) was also submitted to a CC, resulting in 84 fractions analyzed using TLC and brought together according to their respective Rf values. This elucidated Compound **3** (fractions 6–9) and Compound **4** (fractions 38–43). Fractions 40–46 (882 mg) were chromatographed using CC, resulting

in 72 fractions. Then, they were analyzed using TLC and pooled according to their respective Rf values. Posteriorly all compounds isolated, and their, geographic location were made available in the SistematX tool (https://sistematx.ufpb.br) (Scotti et al. 2018).

3.4. Quantitative structure-activity relationship (QSAR) modelling

From the ChEMBL database, 8487 molecules with known activity against *M. tuberculosis* (ChEMBL ID: ChEMBL360) were identified. The molecules were classified using the pMIC (Minimum Inhibitory Concentration) (–log MIC), where the MIC represents the minimum concentration necessary to inhibit the growth of 50% of the bacteria. Compounds with pMIC ≥ 4.7 were considered active, totaling 4181 active molecules. Compounds with pMIC <4.7 were considered inactive, totaling 4306 inactive molecules. Additionally, 68 diterpenes were assessed through virtual screening to identify molecules with potential activity against *M. tuberculosis*. For all structures, SMILES codes were used as input data for Marvin 19.27.0, 2019, ChemAxon (http://www.chemaxon.com). Standardizer software was also used [JChem 19.27.0, 2019; ChemAxon (http://www.chemaxon.com)] to convert the various chemical structures into unique canonical representations. The predictive model was generated in the KNIME Analytics Platform 3.6 software (Berthold et al. 2009), using Random Forest and, through WEKA. The molecular descriptors were calculated using Dragon 7.0 software (Kode srl 2019). The molecules with molecular descriptors ware imported into KNIME, where the data was divided using a "Partitioning" tool set to a stratified sample. This separated the database into training and testing sets, representing 80% and 20% of all compounds, respectively (Monteiro et al. 2020).

3.5. Predicting absorption, distribution, excretion, toxicity properties, hepatic metabolites and ADME approaches

The oral absorption rate (% ABS) was calculated based on the total topological area of the surface (TPSA) using the equation: % ABS = 109 – (TPSA x 0.345). Violations of Lipinski's rule help to understand the oral bioavailability of prototypes to drugs. Thus, it must meet the following requirements, present a molar mass up to 500 Da, up to 10 hydrogen bond acceptors, 5 hydrogen bond donors and logP above 5, and may only

violate one rule (Capecchi et al. 2019). The cytotoxicity risks were calculated with the aid of OSIRIS Data Warrior 5.0 software, assesses the toxicological parameters of mutagenicity (MUT), carcinogenicity (CAR), toxic effect on the reproductive system (ESR) and skin irritability (IRR) for each isolated compound. Additionally, OSIRIS, was used to calculate the molar mass for each structure, as well as the number of hydrogen donors and recipients, partition coefficient (water/octanol) and the polar topological surface area (TPSA) (Sander et al. 2015). The Molecular Discovery MetaSite 6.0 software was used to evaluate the possible transformations undergone by the molecules in the liver. The biotransformed metabolites generated are the submitted to a model for biological activity and cytotoxicity risk prediction through OSIRIS (Baer-Dubowska and Latacz 2018). The ADME approaches were calculated using the web tool SwissADME (Daina et al., 2017).

3.6. Molecular docking

The proteins were obtained from the Protein Data Bank (https://www.rcsb.org/) under PDB ID: 5YHV (aminotransferase) (Saroj and Biswal), 5VRN (InhA) (Xia et al. 2018) together with their respective ligands. The molecular docking study was performed using the Molegro Virtual Docker 6.0.1 software, (Molegro ApS, Aarhus, Denmark) (Monteiro et al. 2020), using the standard software parameters. For coupling, the compounds were inserted in the SDF format and all water and cofactor molecules removed. A template was generated in each cocrystallized inhibitor and the water molecules were removed. Drugs used clinically were selected as controls: cycloserine (Z), ethambutol (E), ethionamide (Eto), Protionamide (Pto), Pyrazinamide (Z), terizidone (Trd). To validate the molecular docking, redocking was performed to account for the root mean square deviation values for each protein; to be considered valid, this value was not greater than 1.5 Å (Dauzhenka et al. 2020).

3.7. Mycobacterial susceptibility assay

The control drugs rifampicin (RIF) and moxifloxacin (MFX) were purchased from Sigma-Aldrich and isoniazid (INH) was obtained from ACROS Organics. *M. tuberculosis* (H37Ra) and *M. smegmatis* (mc^2155) strains were cultured as previously

described (Sidrônio et al. 2021).

The MIC for each compound tested was carried out in 96-well U-bottom polystyrene microplates (Kumari et al. 2018). RIF, MFX and INH solutions were prepared at a concentration of 1 mM in DMSO. The isolated diterpenes were solubilized in DMSO at a concentration of 20 mM and then diluted in Middlebrook 7H9 broth to achieve a concentration of 500 μM. Serial dilutions were performed in 96-well U-bottom microplates maintaining the final concentration of DMSO at 2.5% for all wells. The mycobacterial suspensions were diluted in 7H9 medium at a theoretical optical density (OD 600 nm) of 0.003 and incubated for 7 days at 37 ºC (Muradás et al. 2018). Then, 30 μl of 0.02% resazurin solution was added for analysis of bacterial growth by reducing resazurin. For the evaluation of MIC for strains of *Mycobacterium smegmatis*, the same parameters as above were used, differentiating only the mycobacterial suspensions that were diluted in 7H9 medium at a theoretical optical density (OD 600 nm) of 0.0005, and the incubation period of 24 hours, at 37 ºC (Muradás et al. 2018, Taneja and Tyagi 2007).