RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 11/10/2025.

Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP

Rafael Vieira

Utilização de inteligência artificial e quimioinformática no desenvolvimento de ferramentas computacionais para o estudo da microdiversidade molecular em exsudatos da fermentação de cacau (*Theobroma cacao* L.)

Araraquara, 2023

Rafael Vieira

Utilização de inteligência artificial e quimioinformática no desenvolvimento de ferramentas computacionais para o estudo da microdiversidade molecular em exsudatos da fermentação de cacau (*Theobroma cacao* L.)

Tese de Doutorado apresentada ao Instituto de Química, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de doutor em Química.

Prof. Dr. Ian Castro-Gamboa

Araraquara, 2023

V658u

Vieira, Rafael

Utilização de inteligência artificial e quimioinformática no desenvolvimento de ferramentas computacionais para o estudo da microdiversidade molecular em exsudatos da fermentação de cacau (Theobroma cacao L.) / Rafael Vieira. -- Araraquara, 2023

309 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Química, Araraquara

Orientador: Ian Castro Gamboa

Coorientadora: Kally Alves de Sousa

1. Aprendizado de máquinas. 2. Cacau. 3. Redes neurais (Computação). 4. Dinâmica molecular. 5. Produtos naturais. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Química, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

IMPACTO POTENCIAL DA PESQUISA

Os resultados alcançados nesta pesquisa de doutorado têm potencial em inserir uma série de repercussões importantes no campo da química de produtos naturais e nas ciências ômicas contemporâneas, tais como:

- **Identificação de novos compostos terapêuticos:** A identificação de 10 moléculas distintas em diferentes fases do processo de fermentação do cacau, com potencial atividade contra alvos biomacromoleculares associados a doenças respiratórias, é um avanço significativo. Isso poderia levar ao desenvolvimento de novos tratamentos para doenças como asma e covid-19, que são grandes desafios de saúde global.
- **Automatização e eficiência:** A oferta de produtos em forma de web plataformas inteligentes que fazem uso de inteligência artificial e técnicas computacionais avançadas, como o LUMIOS, Chemistika e CHEIC, auxiliam na automatização do processamento de dados espectrais e na análise de dinâmica molecular. Ferramentas que permitem não apenas economizar tempo e esforço, mas também minimizar o risco de erros humanos, permitindo que os pesquisadores se concentrem em interpretações e aplicações mais avançadas dos dados.
- **Interdisciplinaridade:** A integração de técnicas de química de produtos naturais, quimioinformática, estatística, docagem e dinâmica molecular na análise de matrizes complexas demonstra a natureza interdisciplinar da pesquisa. Contribuindo de maneira fundamental para as ciências ômicas do século XXI, onde a complexidade dos sistemas biológicos (e das matrizes complexas) requerem uma abordagem integrada.
- **Aplicabilidade:** As ferramentas desenvolvidas são altamente aplicáveis a outras áreas de pesquisa em ciências ômicas e química de produtos naturais. Isso poderia facilitar a descoberta de novos compostos bioativos em outras matrizes biológicas complexas, acelerando o desenvolvimento de novos fármacos e tratamentos.

- **Democratização do acesso:** Ao tornar as plataformas gratuitas, permite-se que pesquisadores, independentemente de seus recursos financeiros ou institucionais, tenham acesso a ferramentas avançadas de análise. Isso é especialmente importante para pesquisadores em países em desenvolvimento ou instituições com recursos limitados.
- **Promoção da colaboração:** A disponibilidade de plataformas gratuitas desenvolvidas durante a elaboração desta tese pode facilitar a colaboração entre pesquisadores de diferentes áreas e instituições, uma vez que todos podem acessar e utilizar as mesmas ferramentas. Isso pode levar a uma maior integração e cooperação na comunidade acadêmica.
- **Contribuição para a economia local:** A exploração da cadeia produtiva, por meio dos processos fermentativos, como os do cacau, para a identificação de moléculas com atividade terapêutica, é um exemplo de economia circular. Isso não apenas contribui para a saúde humana, mas também para a sustentabilidade ambiental e econômica de regiões menos favorecidas, como os estados do centro-norte do país.
- **Aceleração da pesquisa:** As ferramentas computacionais podem processar grandes volumes de dados muito mais rapidamente do que seria possível manualmente. Ao disponibilizar essas ferramentas gratuitamente, pode-se acelerar significativamente o progresso da pesquisa em várias áreas.
- **Inovação:** Ao disponibilizar tais plataformas de maneira gratuita, cria-se um ambiente propício para a inovação, pois permite que pesquisadores de diferentes áreas e perspectivas tenham acesso e contribuam para o desenvolvimento e aprimoramento das ferramentas.
- **Educação:** Plataformas gratuitas são recursos valiosos para a educação e treinamento de estudantes e jovens pesquisadores, o quais podem aprender e praticar novas técnicas e métodos de análise sem a barreira de custos associados ao software proprietário.
- **Transparência e reprodutibilidade:** A disponibilidade de plataformas gratuitas pode ajudar a aumentar a transparência e a reprodutibilidade na

pesquisa, pois permite que outros pesquisadores testem e validem os resultados uns dos outros usando as mesmas ferramentas.

Portanto, esta pesquisa não apenas contribui para a química de produtos naturais, fornecendo novos compostos potencialmente terapêuticos e ferramentas para sua identificação e análise, mas também para as ciências ômicas do século XXI, fornecendo uma abordagem interdisciplinar e integrada para a análise de sistemas biológicos complexos, promovendo também a colaboração e a inovação, com intuito de acelerar a pesquisa, melhorar a educação, e aumentar a transparência e a reprodutibilidade na pesquisa.

POTENTIAL IMPACT OF RESEARCH

The results achieved in this doctoral research have the potential to bring about a series of significant repercussions in the field of natural product chemistry and contemporary omic sciences, such as:

• **Identification of new therapeutic compounds:** The identification of 10 distinct molecules at different stages of cocoa fermentation, with potential activity against biomacromolecular targets associated with respiratory diseases, is a significant advancement. This could lead to the development of new treatments for diseases such as asthma and covid-19, which are major global health challenges.

• **Automation and efficiency:** The provision of products in the form of smart web platforms that use artificial intelligence and advanced computational techniques, like LUMIOS, Chemistika, and CHEIC, assist in the automation of spectral data processing and molecular dynamics analysis. Tools that not only save time and effort but also minimize the risk of human errors, allowing researchers to focus on more advanced data interpretations and applications.

• **Interdisciplinarity:** The integration of natural product chemistry techniques, cheminformatics, statistics, docking, and molecular dynamics in the analysis of complex matrices demonstrates the interdisciplinary nature of this research. Contributing fundamentally to the 21st-century omic sciences, where the complexity of biological systems (and of complex matrices) requires an integrated approach.

• **Applicability:** The tools developed are highly applicable to other areas of research in omic sciences and natural product chemistry. This could facilitate the discovery of new bioactive compounds in other complex biological matrices, speeding up the development of new drugs and treatments.

• **Democratizing access:** By making platforms available for free, it allows researchers, regardless of their financial or institutional resources, to access advanced analysis tools. This is especially important for researchers in developing countries or institutions with limited resources.

• **Promoting collaboration:** The availability of free platforms developed during this thesis can facilitate collaboration between researchers from different areas and institutions, as everyone can access and use the same tools. This can lead to greater integration and cooperation in the academic community.

• **Contribution to the local economy:** Exploring the production chain through fermentative processes, such as cocoa's, for identifying molecules with therapeutic activity is an example of a circular economy. This not only contributes to human health but also to the environmental and economic sustainability of less favored regions, like the north-central states of the country.

• **Accelerating research:** Computational tools can process large volumes of data much faster than manually possible. By making these tools available for free, research progress in various areas can be significantly accelerated.

• **Innovation:** By offering such platforms for free, it creates an environment conducive to innovation, as it allows researchers from different fields and perspectives to access and contribute to the development and enhancement of the tools.

• **Education:** Free platforms are valuable resources for educating and training students and young researchers, who can learn and practice new techniques and methods of analysis without the cost barrier associated with proprietary software.

• **Transparency and reproducibility:** The availability of free platforms can help increase transparency and reproducibility in research, as it allows other researchers to test and validate each other's results using the same tools.

Therefore, this research not only contributes to natural product chemistry by providing potentially therapeutic new compounds and tools for their identification and analysis, but also to the 21st-century omic sciences by providing an interdisciplinary and integrated approach to the analysis of complex biological systems, further promoting collaboration and innovation with the intent of accelerating research, improving education, and increasing transparency and reproducibility in research.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Utilização de inteligência artificial e quimioinformática no desenvolvimento de ferramentas computacionais para o estudo da microdiversidade molecular em exsudatos da fermentação de cacau (*Theobroma cacao* L.)"

AUTOR: RAFAEL VIEIRA
ORIENTADOR: IAN CASTRO GAMBOA
COORIENTADORA: KALLY ALVES DE SOUSA

Aprovado como parte das exigências para obtenção do Título de Doutor em Química, pela Comissão Examinadora:

Prof. Dr. IAN CASTRO GAMBOA (Participaçao Presencial)
Departamento de Bioquimica e Quimica Organica / Instituto de Quimica UNESP Araraquara

Prof. Dr. RICARDO ROBERTO DA SILVA (Participaçao Presencial)
Departamento de Física e Química / Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP - Ribeirão Preto

Dr. LUCIANO DA SILVA PINTO (Participaçao Presencial)
Departamento de Química / Centro de Ciências Exatas e de Tecnologia - UFSCAR - São Carlos

Profa. Dra. ERICA REGINA FILLETTI NASCIMENTO (Participaçao Presencial)
Departamento de Engenharia, Física e Matemática / Instituto de Química - UNESP - Araraquara

Profa. Dra. KELLY JOHANA DUSSAN MEDINA (Participaçao Presencial)
Departamento de Engenharia, Fisica e Matematica / Instituto de Quimica - UNESP - Araraquara

Araraquara, 11 de outubro de 2023

IDENTIFICAÇÃO

Nome: Rafael Vieira

Site: www.vieira-rafael.com

Nome em citações bibliográficas: VIEIRA, R.

ORCID: https://orcid.org/0000-0001-9003-3209

Endereço profissional: Rua Prof. Francisco Degni, nº 55, Bairro Quitandinha, Araraquara-SP, CEP: 14800-060

FORMAÇÃO ACADÊMICA/TITULAÇÃO

2019 – 2023 – Doutorado em Química:

Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, São Paulo, Brasil

Título: Relações moleculares da microdiversidade presente em exsudatos da fermentação de cacau (*Theobroma cacao L.*) utilizando *machine learning,* quimioinformática e técnicas de desreplicação.

Orientador: Ian Castro-Gamboa

2014 – 2015 – Mestrado em Química

Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, São Paulo, Brasil

Título: Título: Exploração racional da rede metabólica de *Xylaria* sp. visando a produção de metabólitos de interesse farmacológico, através de ferramentas quimiométricas e técnicas de desreplicação, **Ano de obtenção**: 2015

Orientador: Ian Castro-Gamboa

2008 – 2012 – Graduação em Química

Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, São Paulo, Brasil

FORMAÇÃO COMPLEMENTAR

2021 – 2023:

Especialização em Ciência de Dados e Inteligência Artificial

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – Campus Campinas

Orientador: Prof. Dr. Samuel Botter Martins

2021 – Curso de Curta Duração (virtual):

Amazon Web Service – AWS ACADEMY MACHINE LEARNING (20 horas)

Amazon Web Service – AWS ACADEMY CLOUD FOUNDATIONS (20 horas)

ATUAÇÃO PROFISSIONAL

Período: 16/01/2016 – 31/06/2017

Profissão: Professor e coordenador de curso de Química no Centro Universitário FAEMA – Ariquemes – RO

Período: 17/07/2017 – 31/11/2017

Profissão: Professor efetivo no Instituto Federal de Educação, Ciência e Tecnologia do ACRE (IFAC), Câmpus Tarauacá.

Período: 29/12/2017 – atual

Profissão: Professor efetivo no Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO), Campus Ji-Paraná.

Coordenador de Pós-Graduação vinculado ao Departamento de Pesquisa, Inovação e Pós-Graduação (DEPESP) do *campus* Guajará-Mirim – IFRO (**PORTARIA Nº 29/GJM - CGAB/IFRO, DE 08 DE FEVEREIRO DE 2018)**

PRODUÇÃO BIBLIOGRÁFICA

ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICO:

1. **VIEIRA, RAFAEL**; ALVES DE SOUSA, KALLY; SOUZA DA SILVA, GIVALDO ; HELENA SIQUEIRA SILVA, DULCE ; CASTRO-GAMBOA, IAN . CHEIC: CHEMICAL IMAGE CLASSIFICATOR An intelligent system for identification of volatiles compounds with potential for respiratory diseases using Deep Learning. **EXPERT SYSTEMS WITH APPLICATIONS**. Fator de Impacto (JCR 2022: 8,5), v. 234, p. 121178, **2023**.

2. **VIEIRA, RAFAEL**; DE SOUSA, KALLY ALVES; MONTEIRO, AFIF FELIX; PINTO, LUCIANO SILVA; CASTRO-GAMBOA, IAN. Induction of metabolic variability of the endophytic fungus *Xylaria* sp. by OSMAC approach and experimental design. **ARCHIVES OF MICROBIOLOGY**. Fator de Impacto (JCR 2021: 2,6670, 203, 3025-3022, **2021**.

3. **VIEIRA, R.;** SOUZA, J. M.; SILVA, G. S.; SOUSA, C. O. Identification of Tannins in Amazon Biodiversity Plants: Application Possibilities as a Natural Coagulant. **JOURNAL OF APPLIED OF PHARMACEUTICAL SCIENCE.** , v.5, p.17 - 23, **2018**.

4. HONORATO DE JESUS, JOCIEL; DO CARMO SILVA DE OLIVEIRA, MARIA; MARIA MINETTO BRONDANI, FILOMENA; ROSSI OLIVEIRA LIMA, REGIANE; **VIEIRA, RAFAEL.** PROPRIEDADES FÍSICO-QUÍMICAS DO AMIDO DO CARÁ (*Dioscorea cayennensis*) NATIVO E MODIFICADO POR ACETILAÇÃO. The Journal of Engineering and Exact Sciences. , v.4, p.0429 - 0436, **2018**.

APRESENTAÇÃO DE TRABALHOS

2022 – (Cartagena das Índias – Colômbia):

Apresentação de Poster/Painel no **IV-LAMPS – Latin American Metabolic Profiling Society:** Exploratory analysis of complex matrices of cupuaçu fermented (Theobroma grandiflorum (Willd. ex Spreng.) Schum.): the use of cheminformatics for the bioprospecting of molecules of high added value.

2022 – (Apresentação de palestra (WEBINAR) internacional) organizado pela IV-LAMPS – Latin American Metabolic Profiling Society

Molecular relationship of microdiversity present in cocoa fermentation exudates (*Theobroma cacao* L.) using *Machine Learning* and *Big Data* Technologies

PROJETOS DE PESQUISA ENVOLVIDO (PERÍODO DO DOUTORADO)

2019 – Atual (Pesquisador Responsável) – Projeto Universal – Áreas prioritárias (EDITAL Nº 6/2021/FAPERO-DC)

Análise exploratória das matrizes complexas de fermentados de cupuaçu (Theobroma grandiflorum (Willd. ex Spreng.) Schum.): o uso da quimioinformática para a bioprospecção de moléculas de alto valor agregado

2018 – 2019 (Projetos como Co-orientador):

- Desreplicação metabólica de *Fusarium* sp. e *Aspergillus* sp. cultivados em farelo de cacau (*Theobroma cacao* L.) (Edital no 14/2018/REIT-PROPESP/IFRO, de 10 de maio de 2018) (SEI no 0318471);
- Exploração racional do fermentado de cupuaçu (*Theobroma grandiflorum*) visando a identificação de metabólitos secundários de interesse comercial através de desreplicação (Edital no 14/2018/REIT-PROPESP/IFRO, de 10 de maio de 2018) (SEI no 0318471);
- Exploração racional da espécie *Himatanthus sucuuba* (Spruce) Woodson visando a identificação de metabólitos secundários de interesse comercial (Edital no 48/2018/GJM-CGAB/IFRO, de 22 de junho de 2018) (SEI no 0335263).

PARTICIPAÇÃO EM BANCAS (Últimos 5 anos)

Participação em bancas de curso de Mestrado – 1

Participação em banca de Tiago Teodoro de Lima Souza. Desreplicação do processo fermentativo espontâneo de cacau (*Theobroma cacao* L.) para identificação de bioativos de interesse comercial, **2022**.

Participação em bancas de trabalho de Conclusão de Curso – 4

1 Participação em banca de DANDARA DA SILVA PEREIRA. Avaliação da qualidade físico-química e microbiológica da água consumida no Instituto Federal de Rondônia campus Ji-Paraná, **2019**. (Química) Instituto Federal de Educação Ciência e Tecnologia de Rondônia

2 Participação em banca de STEPHANIE JEDOZ STEIN. Índice de balneabilidade no Rio Machado na área urbana do município de Ji-Paraná - Rondônia**, 2019**. (Química) Instituto Federal de Educação Ciência e Tecnologia de Rondônia

3 Participação em banca de FABIANA DE OLIVEIRA DA SILVA. O uso de Achachairu (*Garcinia humilis*) como Indicador Ácido-Base Natural, **2019** (Química) Instituto Federal de Educação Ciência e Tecnologia de Rondônia.

4 Participação em banca de Danilo Moura Santos. O uso da tecnologia 3D no ensino de química através da confecção de um dominó sobre funções orgânicas, **2018**. (Química) Instituto Federal de Educação Ciência e Tecnologia de Rondônia

ORIENTAÇÕES

Mayara Pacheco Figueiredo. Perfil metabólico da fermentação da amêndoa de cacau (*Theobroma cacao*) utilizando quimiometria e técnicas cromatográficas. **2019**. Curso (Química) - Instituto Federal de Educação Ciência e Tecnologia de Rondônia

PARTICIPAÇÃO EM BANCAS E ORIENTAÇÕES NA CARREIRA:

Orientações concluídas (TCC) – **15**

Participação em banca de trabalhos de conclusão (mestrado) – **1**

Participação em banca (curso de aperfeiçoamento/especialização) – **6**

Participação em banca de trabalhos de conclusão (graduação) – **11**

DEDICATÓRIA

Os filhos que criou trilharam seus caminhos (como dizem que deve ser) ...

Ele não se conformou e andou em nossos calcanhares. Lançou-se em mil direções e disse que no final daquele ano (2019) iria para Rondônia e voltaríamos de carro para Matão. Quando chegássemos, iríamos a São Paulo assistir um jogo do Palmeiras no Allianz Parque.

Toda noite, também sei que ele espreitava nossos antigos quartos para ver se as memórias dormiam direito; se escovamos dentes, se estávamos descobertos...

Ele se fragmentou ainda mais quando estivemos longe, mas no meio de 2019, infelizmente, virou poeira de gente e foi soprado entre nós.

Talvez não teríamos dito um ao outro o quanto nos amávamos (ou tínhamos? Acredito que ao nosso modo, sim), e como pai excessivo que era, não se importou em nenhum momento em pegar as rodovias desse país com seu velho caminhão e nos trazer alimentos quando crianças; ele também nunca pensara em perda e nem permanência; só buscou (ao seu modo simples) nos dar o melhor que pôde.

Não viu seu filho do meio, "*mesmo não sendo médico, virar doutor*" (como ele dizia orgulhosamente aos vizinhos) e, apesar dos percalços da vida, nos mostrou, do seu jeito simples, o caminho da estrada... e eu segui. E ainda sigo.

Se o silêncio da morte é grande, o do nosso coração, pai, é maior ainda...

Esta tese de doutorado é dedicada, *in memoriam,* a Luís Vieira, meu velho pai... Quantas histórias...

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão ao Professor Ian Castro-Gamboa, que não apenas aceitou meu retorno à UNESP após uma empreitada acadêmica muito valiosa no norte do país, mas também pela amizade e por incentivar que eu seguisse por esta linha de pesquisa.

Agradeço aos meus colegas do NUBBE e da pós-graduação, especialmente Givaldo Souza Silva, Tiago Teodoro de Lima Souza, Luciano da Silva Pinto, Camila Cunha, Ana Zanata e Helena Russo, pela camaradagem e apoio contínuo. Sou igualmente grato aos técnicos administrativos e de laboratório, cuja dedicação incansável facilitou grandemente essas análises.

Um agradecimento especial ao Instituto Federal de Rondônia, campus de Ji-Paraná, por permitir minha ausência das atividades acadêmicas por três anos e meio. Sou eternamente grato à Prof. Dra. Kally Alves de Sousa, minha coorientadora e querida amiga, com a esperança de que este seja apenas o começo de muitos outros projetos e artigos colaborativos, e que nossa amizade continue a se fortalecer.

Agradeço também à Fundação de Amparo à Pesquisa de Rondônia (FAPERO) e ao CNPQ pelo suporte financeiro concedido através do edital Universal. Muitos dos resultados apresentados neste trabalho foram possíveis graças a este projeto.

Por fim, agradeço a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho. Cada um de vocês desempenhou um papel crucial na minha jornada, e sou eternamente grato por isso.

RESUMO

Este trabalho está subdividido em seis capítulos. O capítulo 1 tem por objetivo explorar a variabilidade metabólica de produtos de síntese microbiana, provenientes da fermentação natural de sementes de cacau (*Theobroma cacao* L.), na identificação de bioativos de interesse industrial e/ou biotecnológico. Para isso, desenvolveu-se o aplicativo multitarefas LUMIOS, um sistema inteligente que consolida algoritmos para desreplicação, integrando modelos de *machine learning* e abordagens computacionais, o que inclui a docagem molecular, que visam o reconhecimento de feições moleculares em produtos naturais que possam atuar em alvos moleculares associados a doenças respiratórias, como asma e SARS-CoV-2. O capítulo 2 destaca a exploração das misturas complexas do inventário da diversidade metabólica dos exsudatos das sementes fermentadas (e não fermentadas), utilizando o aplicativo LUMIOS, com o intuito de realizar a identificação de sinais oriundos de moléculas de interesse comercial através de ferramentas de desreplicação. O algoritmo LUMIOS efetua comparações com mais de um milhão e duzentos mil espectros de massas, o qual possibilitou a identificação de 13 anotações moleculares, sendo que 10 delas (catequina, trealose, teobromina, procianidina, adenina, indol-3-acetamida, ácido ftálico, anidrido ftálico, fenilalanina e tirosina) apresentaram potenciais para atuarem em alvos de doenças respiratórias. O capítulo 3 apresenta o desenvolvimento e a testagem do aplicativo Chemistika, que, aliado às anotações fornecidas pelo LUMIOS, permite a automatização e tratamento de dados do planejamento de misturas do tipo Simplex-Lattice, do tipo 3x3. Os capítulos 4 e 5 apresentam a utilização do Chemistika para construir modelos que possam prever a intensidade relativa de cada anotação oriunda nas matrizes complexas de cacau e explorar a variabilidade metabólica nas diferentes fases do complexo processo fermentativo. Por fim, o capítulo 6 explora cada anotação molecular apontada pelo LUMIOS à luz da dinâmica molecular (utilizando o algoritmo Gromacs), em um estudo de trajetória de 100 nanossegundos, e faz uso da plataforma CHEIC para analisar os resultados ofertados pelo Gromacs. Os produtos computacionais desenvolvidos neste

trabalho, LUMIOS, Chemistika e CHEIC, representam avanços significativos na exploração de matrizes de produtos naturais. Essas ferramentas não apenas automatizam e agilizam o processo de análise, mas também proporcionam uma compreensão mais profunda das complexas interações moleculares presentes nos produtos naturais. A capacidade de identificar rapidamente moléculas de interesse comercial e biotecnológico, prever a intensidade relativa de anotações moleculares e explorar a dinâmica molecular de compostos promissores tem o potencial de acelerar a descoberta de novos bioativos e otimizar o processo de desenvolvimento de novos produtos. Além disso, ao facilitar a exploração racional da microdiversidade presente em sementes de cacau, essas ferramentas podem contribuir para a valorização deste recurso natural e para o desenvolvimento de iniciativas biotecnológicas inovadoras.

Palavras-chave: inteligência artificial; softwares multitarefas; planejamento de misturas; COVID-19; docking; dinâmica molecular

ABSTRACT

This work is divided into six chapters. Chapter 1 aims to explore the metabolic variability of microbial synthesis products, derived from the natural fermentation of cocoa seeds (Theobroma cacao L.), in the identification of bioactive compounds of industrial and/or biotechnological interest. For this purpose, the multi-tasking application LUMIOS was developed, an intelligent system that consolidates algorithms for dereplication, integrating machine learning models and computational approaches, which includes molecular docking, aimed at recognizing molecular features in natural products that may act on molecular targets associated with respiratory diseases, such as asthma and SARS-CoV-2. Chapter 2 highlights the exploration of the complex mixtures of the metabolic diversity inventory of the exudates of fermented (and non-fermented) seeds, using the LUMIOS application, with the aim of identifying signals originating from molecules of commercial interest through dereplication tools. The LUMIOS algorithm performs comparisons with more than one million two hundred thousand mass spectra, which enabled the identification of 13 molecular annotations, 10 of which (catechin, trehalose, theobromine, procyanidin, adenine, indole-3-acetamide, phthalic acid, phthalic anhydride, phenylalanine, and tyrosine) showed potential to act on targets of respiratory diseases. Chapter 3 presents the development and testing of the Chemistika application, which, together with the annotations provided by LUMIOS, allows the automation and data processing of the Simplex-Lattice mixture design, 3x3 type. Chapters 4 and 5 present the use of Chemistika to construct models that can predict the relative intensity of each annotation originating in the complex matrices of cocoa and explore the metabolic variability in the different phases of the complex fermentation process. Finally, chapter 6 explores each molecular annotation pointed out by LUMIOS in the light of molecular dynamics (using the Gromacs algorithm), in a 100 nanosecond trajectory study, and makes use of the CHEIC platform to analyze the results provided by Gromacs. The computational products developed in this work, LUMIOS, Chemistika, and CHEIC, represent significant advances in the exploration of natural product matrices. These tools not

only automate and speed up the analysis process but also provide a deeper understanding of the complex molecular interactions present in natural products. The ability to quickly identify molecules of commercial and biotechnological interest, predict the relative intensity of molecular annotations, and explore the molecular dynamics of promising compounds has the potential to accelerate the discovery of new bioactives and optimize the process of developing new products. Furthermore, by facilitating the rational exploration of the microdiversity present in cocoa seeds, these tools can contribute to the valorization of this natural resource and the development of innovative biotechnological initiatives.

Keywords: artificial intelligence; multitasking software; mixture design; COVID-19; docking; molecular dynamics.

LISTA DE FIGURAS

FIGURAS – MATERIAIS SUPLEMENTARES

LISTA DE TABELAS

LISTA DE TABELAS SUPLEMENTARES

SUMÁRIO

INTRODUÇÃO GERAL

Esta pesquisa envolveu a cadeia produtiva de cacau, principalmente de base familiar, inserida na mesorregião do leste rondoniense, e relacionou o emprego (e desenvolvimento) de técnicas e ferramentas inovadoras para identificar produtos oriundos do processo fermentativo de sementes de cacau. Este trabalho é fruto de uma colaboração nacional entre o Instituto Federal de Rondônia – IFRO (Ji-Paraná) e Instituto de Química da UNESP (Araraquara), com o propósito de criar uma sinergia de ferramentas científico-tecnológicas para o desenvolvimento de processos biotecnológicos de importância regional e/ou internacional.

No escopo de tais alternativas inovadoras foram empregadas ferramentas da Quimioinformática, da Ciência de Dados e da Inteligência Artificial, por meio do aprendizado de máquina (*Machine Learning)* e do aprendizado profundo (*Deep Learning)*, para resolver problemas no campo da química, como recuperação e extração de informações químicas, pesquisa de banco de dados, mineração de espaços químicos moleculares e criação de modelos capazes de prever faixas mais amplas de padrões de fármaco-similaridade, uma vez que a rápida explosão de *Big Data* de dados químicos e a crescente necessidade da redução de tempo para descoberta de moléculas-fármacos são recorrentes, abordagens computacionais se tornaram ferramenta indispensável para extrair informações de bancos e desenvolver medicamentos com propriedades biológicas importantes.

O *Machine Learning* e o *Deep Learning* são atualmente um dos tópicos mais importantes e em rápida evolução na descoberta de medicamentos auxiliados por computador. Em contraste com os modelos físicos que dependem de equações físicas explícitas, como Química Quântica ou simulações de dinâmica molecular, as abordagens de aprendizagem de máquina usam algoritmos de reconhecimento de padrões para discernir relações matemáticas entre observações empíricas de pequenas moléculas com o intuito de extrapolá-las para prever propriedades químicas, biológicas e físicas de novos compostos. Além disso, a mineração matemática de entidades químicas permite a derivação de uma constelação de descritores, que são empacotados como impressões digitais químicas em uma

variedade de modelos de aprendizado de máquina podendo explorar de maneira eficiente a variabilidade metabólica de um extrato bruto, prevendo padrões de estruturas que possam atuar em alvos biomacromoleculares de doenças diversas.

Desta forma, essa tese será dividida em capítulos, que em conjunto, abordarão: **(i)** curadoria de dados e desreplicação dos exsudatos de sementes fermentadas de cacau por meio de delineamento experimental do tipo misturas; **(ii)** análise de espaço químico, visualização, navegação e comparação das moléculas prospectadas nestes exsudatos, buscando anotações moleculares significativas; **(iii)** previsão de bioatividade das moléculas anotadas e que foram sinalizadas como promissoras por meio de inteligência artificial, docagem e dinâmica molecular **(v)** abordagens computacionais para automatização das análises químicas e desenvolvimento de softwares.

Tal abordagem enfatiza a importância do cacau não apenas na produção de chocolate, mas também como formas alternativas de obtenção de moléculas de interesse comercial para indústrias cosméticas, alimentícias e farmacêuticas.

1 REVISÃO DA LITERATURA

1.1 Cacau.

O cultivo, comercialização e industrialização do cacau (*Theobroma cacao* L.*)* e seus derivados têm apresentado, no decorrer dos anos, importante papel socioeconômico no cenário brasileiro e latino-americano (FRANZEN; BORGERHOFF MULDER, 2007).

A qualidade do chocolate, principal produto obtido do cacau, depende de uma grande variedade de fatores ambientais, agronômicos e tecnológicos (COOPER et al., 2008). Porém, este trabalho vai além dos potenciais de produção de chocolate, ele sinaliza o uso do cacau para indústria farmacêutica, alimentícia e cosmética, no sentido de obtenção de moléculas de alto valor agregado.

Dentre esses fatores, sabe-se que os micro-organismos, presentes na fermentação espontânea das sementes do cacau, desempenham função essencial no desenvolvimento dos metabólitos (MORENO-ZAMBRANO et al., 2018). Considerando que o processo de fermentação e secagem é realizado ainda nas fazendas, sem qualquer controle de processo, uma porcentagem significativa das sementes não sofre as alterações necessárias (principalmente a acidificação do pH e aumento da temperatura) para que as reações enzimáticas se processem de forma satisfatória. Uma possibilidade de remediar este problema é o acompanhamento e intervenção, principalmente no processo de fermentação, objetivando caracterizar os compostos aromáticos, enzimas e melhores condições de processo para melhor uniformizar e aumentar a qualidade das amêndoas de cacau produzidas (CASTRO-ALAYO et al., 2019).

A fermentação do cacau é um processo microbiológico espontâneo, no qual os micro-organismos metabolizam os açúcares fermentescíveis presentes na polpa, em ácido lático e etanol, o qual posteriormente, é oxidado a ácido acético através de reação exotérmica que envolve a atuação de bactérias acéticas. A polpa de cacau é um substrato rico para desenvolvimento microbiano, consistindo em 82-87% de água, 10-15% de açúcar, 2-3% de pentosanas, 1-3% de ácido cítrico e 1-

1,5% de pectina. Proteínas, aminoácidos, vitaminas e minerais também estão presentes (KONGOR et al., 2016a).

O ácido acético e o etanol, produtos do metabolismo microbiano, penetram na semente e em combinação com a ação do calor eliminam a capacidade germinativa do embrião quebrando as paredes celulares da semente. Estas alterações induzem reações bioquímicas dentro da amêndoa, gerando os precursores químicos do sabor e cor do chocolate (SANTANDER MUÑOZ et al., 2020).

O tempo requerido para a fermentação das sementes de cacau é variável. Para a ocorrência das principais reações, que levam à formação dos principais precursores moleculares, as sementes de cacaueiro do grupo Forastero, tipo predominante em todo o mundo, inclusive no Brasil, deve ser geralmente fermentado por períodos superiores a cinco dias (BRUNETTO et al., 2020a).

Muitos micro-organismos fermentadores são provenientes das mãos dos trabalhadores que manipulam os frutos durante os procedimentos para o rompimento das cascas. Ademais, os micro-organismos são oriundos dos cestos utilizados para transporte das sementes e da mucilagem seca, presente nas caixas, remanescente de fermentações anteriores (PUERARI; MAGALHÃES; SCHWAN, 2012; SCHWAN, 1998).

1.2 Quimiometria e desreplicação molecular.

Inúmeras tecnologias têm sido desenvolvidas para explorar e identificar metabólitos secundários, de interesse industrial e farmacológico, produzidos durante processos fermentativos. Nesta perspectiva, destaca-se a quimiometria, que faz uso de ferramentas estatísticas, e a metabolômica, com a abordagem de desreplicação, que visa detectar novos compostos sem a necessidade de isolamento (HILLMAN; READNOUR; SOLOMON, 2017).

A quimiometria é uma área da ciência que utiliza conhecimentos de matemática e estatística para a identificação de informações relevantes de um problema em estudo, facilitando a obtenção de informações (KJELDAHL; BRO,

2010). Durante o processamento dos dados, pode-se realizar uma análise exploratória dos dados químicos fazendo uma varredura nos cromatogramas e nos espectros obtidos, buscando evidências de sinais de moléculas de alto valor agregado.

Adicionalmente, buscando realizar induções planejadas, a quimiometria contribuiu nessa investigação ao criar um delineamento experimental com o intuito de otimizar as condições de extração dos metabólitos oriundos da fermentação das sementes de cacau. Esse tipo de desenho experimental, conhecido em inglês por *Design of Experiments* (DoE), agrega o planejamento fatorial e possibilita o estudo das condições experimentais ideais de um dado problema (POLITIS et al., 2017). Para se obter sucesso na utilização desta técnica, é necessário estimar todos os possíveis fatores que podem impactar na fermentação das sementes de cacau. O planejamento experimental permite investigar, de forma robusta e econômica, os efeitos de vários fatores sobre as respostas de interesse, com o objetivo deste tipo de planejamento é encontrar a combinação ideal dos componentes que maximize a resposta desejada (CURTIS et al., 2022).

O trabalho de identificação, reconhecimento e elucidação estrutural é uma tarefa árdua, fazendo-se uso de técnicas modernas baseadas no princípio da química verde e o uso de simuladores *in silico* (computadores de alta performance) tem atraído atenção de renomados grupos de pesquisa em todo o mundo e tal abordagem tem sido conduzida de forma crescente no Brasil. Recentemente, a desreplicação pode ser auxiliada pela técnica computacionais, que faz relações moleculares com as fragmentações propiciadas pelo espectrômetro de massas (TARTAGLIONE et al., 2023). Essa abordagem metodológica pode ser destinada a distinguir, em matrizes complexas, os compostos já conhecidos daqueles ainda desconhecidos e, que possivelmente apresentem interesse de exploração (KIND; FIEHN, 2017).

Para tal, diversas técnicas hifenadas de separação e detecção são utilizadas, nas quais destacam-se a Cromatografia Gasosa Acoplada à Espectrometria de Massas (CG-EM) e Cromatografia Líquida de Alta Eficiência Acoplada à Espectrometria de Massas (CLAE-EM), bioensaios e análises que permitam a

comparação dos conjuntos de dados obtidos através do uso de base de dados (JOUANEH et al., 2022; QIN et al., 2023).

Tais técnicas, auxiliam na busca de moléculas bioativas já descritas na literatura e disponíveis em bases de dados, possibilitando correlacionar os metabólitos com a cocobiota presente nas sementes fermentadas de cacau fazendo uso de abordagens de ciência de dados, atreladas à quimioinformática, química computacional e inteligência artificial para reconhecer feições moleculares de compostos que possam atuar em alvos de doenças diversas (que neste trabalho, se direcionará à busca de moléculas capazes de modular alvos biomacromoleculares de doenças respiratórias, como bronquite, asma e SARS-CoV-2).

1.3 Abordagens computacionais (Ciência de Dados e Inteligência Artificial).

Para um maior interesse acadêmico e farmacológico, a versatilidade do uso de produtos naturais (PN) em diferentes áreas, como polímeros, suplementos alimentares, agricultura e cosméticos, impulsionou o aumento no número de bancos de dados moleculares abertos e restritos (comerciais) (AHMED et al., 2010; CROTEAU et al., 2000; KULKARNI VISHAKHA; BUTTE KISHOR; RATHOD SUDHA, 2012; SPARKS et al., 2019) e informações químicas agregadas de vários organismos, biomas, doenças específicas e usos tradicionais (SOROKINA et al., 2021).

Um banco de dados químico pode ser definido como uma coleção de moléculas que contém informações sobre compostos e seus descritores químicos, bem como reatividade e diversos recursos biológicos (KOULOURIDI et al., 2019a). Esses bancos de dados são alimentados pela análise de artigos científicos que contêm resultados do processo de isolamento e elucidação estrutural de produtos naturais. Algumas plataformas de banco de dados são sistematicamente analisadas e revisadas, visando racionalizar e organizar as informações disponíveis sobre moléculas orgânicas, às vezes não publicadas, de diferentes origens biológicas.

Informações como estrutura, propriedades biológicas, origem e localização geográfica são inseridas manualmente nos bancos de dados, enquanto as propriedades moleculares e o nome IUPAC, bem como a geração de espectros de ressonância magnética nuclear, são gerados automaticamente (PILON et al., 2017). Assim, a necessidade de armazenar, gerenciar e processar essas informações criou um "Big Data" que contempla um espaço químico diverso (SALDIVAR-GONZALEZ et al., 2018), que pode ser explorado por diferentes métodos *in silico*, permeando ferramentas da inteligência artificial, da quimioinformática, da docagem e da dinâmica molecular.

1.4 Machine learning – ML

A Inteligência Artificial (IA) tem se estabelecido como uma das tecnologias mais influentes do século XXI (ÖZDEMIR; HEKIM, 2018) e o *Machine Learning*, um subconjunto essencial da IA, desempenha um papel fundamental nessa revolução tecnológica (SARKER, 2021). À medida que o mundo se torna cada vez mais dependente de dados, a capacidade de extrair informações valiosas desses dados se torna uma habilidade crítica (QIU et al., 2016).

Machine Learning não é apenas uma ferramenta, mas uma disciplina que revolucionou a maneira como as máquinas compreendem e interpretam informações (SOORI; AREZOO; DASTRES, 2023). Ao invés de depender de regras de programação rígidas, os algoritmos de *Machine Learning* aprendem com dados, refinam suas respostas e se adaptam a novos cenários (WOSCHANK; RAUCH; ZSIFKOVITS, 2020). Isso abre caminho para uma ampla gama de aplicações que vão desde a automação de tarefas rotineiras até a solução de desafios complexos em áreas como medicina, finanças, e indústria (ÇELIK; ALTUNAYDIN, 2018).

Dentro do campo do *Machine Learning*, existem três grandes áreas que constituem seus pilares fundamentais (DAS; DEY; ROY, 2015):

- **Aprendizado supervisionado:** onde um modelo é treinado em dados rotulados, é amplamente utilizado em tarefas como reconhecimento de fala, classificação de imagens e diagnóstico médico.

- **Aprendizado não supervisionado:** O aprendizado não supervisionado, que explora a estrutura latente de dados não rotulados, desempenha um papel crucial em tarefas como segmentação de mercado, análise de redes sociais e detecção de anomalias.
- **Aprendizado por reforço:** O aprendizado por reforço é fundamental em domínios como jogos, robótica e controle de processos. É uma abordagem para a aprendizagem de máquina em que um agente interage com seu ambiente e aprende a tomar ações para maximizar alguma forma de recompensa cumulativa (DAYAN; NIV, 2008).

Em suma, o campo do aprendizado de máquina, ou *machine learning*, está revolucionando a maneira de interagir com a tecnologia e solucionar problemas em diversas áreas. À medida que a pesquisa e a aplicação prática continuam a avançar, pode-se esperar que o aprendizado de máquina desempenhe um papel cada vez mais crucial na sociedade, impulsionando a automação, a tomada de decisões inteligentes e a inovação em todos os setores (RUDIN; WAGSTAFF, 2014). A jornada do *machine learning* está longe de ser concluída, e seu potencial quase que ilimitado promete um futuro empolgante e repleto de possibilidades, sobretudo, na área das ciências.

1.5 Deep Learning – DL

Deep learning, em português "aprendizado profundo," é uma subárea do aprendizado de máquina (*machine learning*) que se concentra na criação e treinamento de redes neurais artificiais profundas para realizar tarefas complexas de análise de dados e tomada de decisão (LECUN; BENGIO; HINTON, 2015). O termo "profundo" refere-se ao fato de que essas redes neurais são compostas por múltiplas camadas de unidades de processamento, conhecidas como neurônios artificiais (MIN; LEE; YOON, 2017).

A principal característica do *deep learning* é a capacidade de aprender automaticamente representações de dados de nível hierárquico, o que permite a

extração de características complexas e abstratas dos dados de entrada (ELHARROUSS et al., 2022). Isso é especialmente útil em tarefas de visão computacional (DEL CAMPO; CARLSON; MANNINGER, 2021), processamento de linguagem natural (KANG et al., 2020), reconhecimento de padrões (BAI et al., 2021), entre outras. Essas redes são treinadas usando algoritmos de otimização para minimizar o erro entre as previsões do modelo e os rótulos (labels) verdadeiros dos dados de treinamento (SUN, 2020).

1.5.1. Arquitetura das redes neurais convolucionais – CNNs

As CNNs são compostas por camadas convolucionais que aplicam operações de convolução para extrair características relevantes dos dados de entrada. Essas camadas são seguidas por camadas de *pooling* para reduzir a dimensionalidade e camadas totalmente conectadas para realizar a classificação ou regressão (BHATT et al., 2021). A arquitetura típica de uma CNN inclui:

- **Camadas de Convolução**: Essas camadas aplicam filtros (*kernels*) a regiões locais da entrada para extrair características relevantes. Cada filtro é aprendido durante o treinamento para detectar padrões específicos (YAMASHITA et al., 2018).
- **Camadas de *Pooling***: As camadas de *pooling* reduzem o tamanho espacial da representação, mantendo as características mais importantes. Isso ajuda a tornar a rede mais eficiente e robusta (SINGH; RAJ; NAMBOODIRI, 2020).
- **Camadas Totalmente Conectadas**: No final da rede, camadas totalmente conectadas são usadas para realizar a classificação ou regressão com base nas características extraídas.

Uma das razões pelas quais o *Deep Learning* se tornou tão poderoso é o uso de conjuntos de dados grandes e avanços em hardware de GPU (*graphics processing units*), que aceleram o treinamento de redes neurais profundas (GAWEHN et al., 2018), propiciando o contínuo desenvolvimento da área, com

avanços constantes em arquiteturas de rede, algoritmos de treinamento e aplicações práticas.

1.6 Inteligência artificial em química

A inteligência artificial, por meio de algoritmos de *machine* e *deep learning*, tem moldado inúmeras áreas do conhecimento, inclusive a química. Suas aplicações são vastas e seu potencial é ilimitado. Em particular, na química, o *Machine Learning* tem desempenhado um papel crucial na aceleração da descoberta de novos compostos (MEUWLY, 2021), na previsão de propriedades moleculares (COVA; PAIS, 2019), na otimização de processos químicos (HORWOOD; NOUTAHI, 2020) e na identificação de padrões em grandes conjuntos de dados experimentais (CHOI et al., 2021).

Uma das contribuições mais significativas do *Machine Learning* na química é a capacidade de projetar moléculas com propriedades específicas, o que é essencial no desenvolvimento de novos medicamentos e materiais (CHAN et al., 2019). Algoritmos de *Machine Learning* podem analisar vastos bancos de dados de compostos químicos e identificar estruturas promissoras que atendem a critérios específicos, economizando tempo e recursos no processo de pesquisa, tornando-a mais eficiente e precisa (PAUL et al., 2021).

À medida que a colaboração entre cientistas e especialistas em *Machine Learning* continua a crescer, é provável que existam avanços ainda mais notáveis na interseção entre a química e o *Machine Learning*. Assim, a área em questão tem experimentado uma significativa e crescente importância nos últimos anos, emergindo como um campo de pesquisa de vanguarda. As pesquisas avançadas realizadas nesse domínio refletem seu impacto duradouro e seu potencial para moldar o futuro de inúmeras esferas da sociedade, sobretudo, na área das ciências. À medida que se continua a explorar as inúmeras implicações e desdobramentos desse campo em constante evolução, pode-se antecipar avanços significativos que irão impactar positivamente uma ampla gama de setores e disciplinas.

1.7 Plataformas web.

A revolução digital no campo da ciência de dados tem transformado várias disciplinas, incluindo a química (STENTA, 2021). Uma consequência direta desta revolução é o acúmulo massivo de dados químicos, consolidados em diversos bancos de dados moleculares disponíveis publicamente ou privadamente. Este acúmulo de dados impulsionou a construção de plataformas web especializadas, projetadas para facilitar e automatizar análises de dados químicos complexos (LI MANNI et al., 2023).

Essas plataformas não apenas tornam o acesso e a análise de dados mais eficientes, mas também facilitam a colaboração entre pesquisadores de diferentes partes do mundo, promovendo assim o avanço da ciência de forma mais globalizada (BATTISTELLA; NONINO, 2012). Além disso, a automação de análises de dados químicos por meio dessas plataformas ajuda a minimizar erros humanos e permite que os pesquisadores se concentrem em aspectos mais criativos e interpretativos de seus estudos (SCHMIDT; LIPSON, 2009).

A disponibilidade dessas plataformas tem implicações significativas para a pesquisa e desenvolvimento em química de produtos naturais, química medicinal, ciências ômicas, entre outras áreas. Elas facilitam a identificação de novos compostos bioativos, a predição de suas propriedades e atividades, e a análise de seus mecanismos de ação. Tudo isso contribui para a aceleração do processo de descoberta de novos fármacos e outros produtos de interesse (LIU et al., 2019) (Gonzalez et al., 2020).

Neste contexto, esta pesquisa de doutorado tem como produto a elaboração de três inovadores aplicativos disponíveis em plataforma web: LUMIOS, Chemistika e CHEIC. Estes foram desenvolvidos com o objetivo de automatizar de maneira inteligente as análises exploratórias dos dados químicos de matrizes complexas de cacau, em suas diversas fases de fermentação. Tais aplicativos não só facilitam e diminuem significativamente o tempo necessário para a análise de dados, mas também integra tecnologias avançadas como inteligência artificial, desreplicação, planejamento experimental, docagem e dinâmica molecular. Dessa forma, os

aplicativos desenvolvidos não apenas otimizam a eficiência do processo de análise, mas também potencializam a qualidade e a relevância dos dados obtidos, contribuindo significativamente para um melhor entendimento das matrizes complexas oriundas durante a fermentação do cacau. Este avanço, por sua vez, tem o potencial de descoberta de novas aplicações para os compostos presentes no cacau.

4 CONCLUSÃO

Tais análises apresentadas e discutidas neste capítulo teve por objetivo realizar uma varredura entre possíveis anotações moleculares de matrizes complexas de cacau, analisando-as como ligantes em potencial para modular alvos bioquímicos. Tal abordagem não apenas proporcionou uma compreensão mais profunda das interações moleculares que tais moléculas possam efetuar com sítios ativos das proteínas associadas a doenças respiratórias, mas também pavimentou o caminho para futuras investigações que poderão ter implicações significativas em áreas como o desenvolvimento de fármacos e a biologia estrutural.

Ao identificar e compreender essas peculiaridades, pode-se preparar estratégias mais inteligentes e rápidas de design de ligantes e buscar candidatos potencialmente mais eficazes para a interação com bioreceptores desta natureza, o que pode ter amplas implicações no campo do desenvolvimento de fármaco.

Em conclusão, apesar de todos as moléculas aqui discutidas serem potenciais para atuar em doenças respiratórias, destacam-se, especialmente, a catequina por formar um complexo altamente estável e específicos com 3 dos 4 receptores propostos (6VVU, 4DD8 e 7P2G). A trealose, com interação geral mais fraca, sugeriu associações mais transitórias e flexíveis, mas estatisticamente tão consistentes quanto aos ligantes padrões das proteínas 6VVU e 4DD8 (e também da catequina). E o ácido ftálico, que apresentou afinidade pela proteína 6VVU.

Estas descobertas propiciadas por análises computacionais robustas, como as de dinâmica molecular, coletivamente sugerem que, embora catequina, trealose e ácido ftálico possam associarem-se ao mesmo receptor, a natureza de suas interações moleculares varia significativamente, modulando o receptor de maneiras distintas, mas resultando em respostas bioquímicas similares.

REFERÊNCIAS

ABDEL-KAREEM, M. M.; RASMEY, A. M.; ZOHRI, A. A. The action mechanism and biocontrol potentiality of novel isolates of Saccharomyces cerevisiae against the aflatoxigenic Aspergillus flavus. **Letters in applied microbiology**, v. 68, n. 2, p. 104–111, 2019.

AFOAKWA, E. O. et al. Chemical composition and physical quality characteristics of Ghanaian cocoa beans as affected by pulp pre-conditioning and fermentation. **Journal of Food Science and Technology**, v. 50, n. 6, p. 1097–1105, dez. 2013.

AGYIRIFO, D. S. et al. Metagenomics analysis of cocoa bean fermentation microbiome identifying species diversity and putative functional capabilities. **Heliyon**, v. 5, n. 7, 2019.

AHMED, J. et al. SuperSweet—a resource on natural and artificial sweetening agents. **Nucleic acids research**, v. 39, n. suppl_1, p. D377–D382, 2010.

AKSIMENTIEV, A.; SCHULTEN, K. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. **Biophysical Journal**, v. 88, n. 6, p. 3745–3761, 2005.

ALIN, A. Minitab. **Wiley interdisciplinary reviews: computational statistics**, v. 2, n. 6, p. 723–727, 2010.

ANDÚJAR, I. et al. **Cocoa polyphenols and their potential benefits for human health**. **Oxidative Medicine and Cellular Longevity**, 2012.

APROTOSOAIE, A. et al. The Cardiovascular Effects of Cocoa Polyphenols—An Overview. **Diseases**, v. 4, n. 4, p. 39, dez. 2016.

ARBORETTI, R. et al. **Design of Experiments and machine learning for product innovation: A systematic literature review**. **Quality and Reliability Engineering International.** John Wiley and Sons Ltd, 1 mar. 2022.

ARTRITH, N. et al. Best practices in machine learning for chemistry. **Nature chemistry**, v. 13, n. 6, p. 505–508, 2021.

ATANASOV, A. G. et al. Natural products in drug discovery: advances and opportunities. **Nature reviews Drug discovery**, v. 20, n. 3, p. 200–216, 2021.

AZCARATE, S. M.; PINTO, L.; GOICOECHEA, H. C. **Applications of mixture experiments for response surface methodology implementation in analytical methods development**. **Journal of Chemometrics.** John Wiley and Sons Ltd, dez. 2020.

BACH, E.; SCHYMANSKI, E. L.; ROUSU, J. Joint structural annotation of small molecules using liquid chromatography retention order and tandem mass spectrometry data. **Nature Machine Intelligence**, v. 4, n. 12, p. 1224–1237, 1 dez. 2022.

BAI, X. et al. **Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments**. **Pattern Recognition.** Elsevier Ltd, dez. 2021.

BAJORATH, J. Integration of virtual and high-throughput screening. **Nature Reviews Drug Discovery**, v. 1, n. 11, p. 882–894, 2002.

BALABAN, A. T. Chemical graphs. **Theoretica Chimica Acta**, v. 53, n. 4, p. 355–375, 1979.

BALENTIC, J. P. et al. Cocoa shell: A by-product with great potential for wide application. **Molecules**, v. 23, n.6, 2018.

BARRIL, X. et al. How accurate can molecular dynamics/linear response and Poisson-Boltzmann/solvent accessible surface calculations be for predicting relative binding affinities? Acetylcholinesterase huprine inhibitors as a test case. **Theoretical Chemistry Accounts**, v. 106, n. 1–2, p. 2–9, jun. 2001.

BART-PLANGE, A.; BARYEH, E. A. The physical properties of Category B cocoa beans. **Journal of Food Engineering**, v. 60, n. 3, p. 219–227, dez. 2003.

BATTISTELLA, C.; NONINO, F. Open innovation web-based platforms: The impact of different forms of motivation on collaboration. **Innovation: Management, Policy and Practice**, v. 14, n. 4, p. 557–575, 2012.

BAYADA, D. M.; HAMERSMA, H.; VAN GEERESTEIN, V. J. Molecular diversity and representativity in chemical databases. **Journal of chemical information and computer sciences**, v. 39, n. 1, p. 1–10, 1999.

BELWAL, T. et al. Bioactive Compounds from Cocoa Husk: Extraction, Analysis and Applications in Food Production Chain. **Foods**, v.11, n.6, p.798, mar. 2022.

BENDER, A. et al. “Bayes affinity fingerprints” improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? **Journal of chemical information and modeling**, v. 46, n. 6, p. 2445–2456, 2006.

BERRY, N. M. et al. Impact of cocoa flavanol consumption on blood pressure responsiveness to exercise. **British Journal of Nutrition**, v. 103, n. 10, p. 1480–1484, 2010.

BHATT, D. et al. **CNN variants for computer vision: History, architecture, application, challenges and future scope**. **Electronics**, v.10, n. 20, p. 2470, out. 2021.

BOLL, M. et al. Microbial degradation of phthalates: biochemistry and environmental implications. **Environmental Microbiology Reports**, v. 12, n. 1, p. 3-15, fev. 2020.

BOOZARI, M.; HOSSEINZADEH, H. Natural products for COVID-19 prevention and treatment regarding to previous coronavirus infections and novel studies. **Phytotherapy Research**, v. 35, n. 2, p. 864–876, 2021.

BRUNETTO, M. DEL R. et al. The effect of fermentation and roasting on free amino acids profile in Criollo cocoa (Theobroma cacao L.) grown in Venezuela. **Brazilian Journal of Food Technology**, v. 23, 2020.

BÜHLMANN, P.; YU, B. Discussion of "Additive logistic regression: A statistical view," by J. Friedman, T. Hastie and R. Tibshirani. **Ann. Statist**, v. 28, p. 377–386, 2000.

BUITRAGO-LOPEZ, A. et al. Chocolate consumption and cardiometabolic disorders: Systematic review and meta-analysis. **BMJ (Online)**, v. 343, n. 7825, 1 out. 2011.

CÁDIZ-GURREA, M. L. et al. Isolation, comprehensive characterization and antioxidant activities of Theobroma cacao extract. **Journal of Functional Foods**, v. 10, p. 485–498, 2014.

CAMANDOLA, S.; PLICK, N.; MATTSON, M. P. Impact of Coffee and Cacao Purine Metabolites on Neuroplasticity and Neurodegenerative Disease. **Neurochemical Research**, v. 44, n. 1, p. 214–227, 15 jan. 2019.

CAMU, N. et al. Influence of turning and environmental contamination on the dynamics of populations of lactic acid and acetic acid bacteria involved in spontaneous cocoa bean heap fermentation in Ghana. **Applied and Environmental Microbiology**, v. 74, n. 1, p. 86–98, jan. 2008.

CASTRO-ALAYO, E. M. et al. Formation of aromatic compounds precursors during fermentation of Criollo and Forastero cocoa. **Heliyon**, v. 5, n. 1, 2019.

ÇELIK, Ö.; ALTUNAYDIN, S. S. A Research on Machine Learning Methods and Its Applications. **Journal of Educational Technology & Online Learning**, v. 1, n. 3, p. 25–40, 2018.

CHAN, H. C. S. et al. Advancing Drug Discovery via Artificial Intelligence. **Trends in Pharmacological** Sciences, v. 40, n. 8, p. 592-604, ago. 2019.

CHAPMAN, A. G.; ATKINSON, D. E. Adenine nucleotide concentrations and turnover rates. Their correlation with biological activity in bacteria and yeast. **Advances in microbial physiology**, v. 15, p. 253–306, 1977.

CHASALOW, S. D.; BRAND, R. J. Algorithm AS 299: Generation of Simplex Lattice Points. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 44, n. 4, p. 534-545, 1995.

CHEN, D. et al. **GREEN TEA AND TEA POLYPHENOLS IN CANCER PREVENTIONFrontiers in Bioscience**. [s.l: s.n.].

CHOI, J. et al. Exploring the chemical space of protein–protein interaction inhibitors through machine learning. **Scientific Reports**, v. 11, n. 1, dez. 2021.

CHOURASIA, M. et al. EGCG, a green tea catechin, as a potential therapeutic agent for symptomatic and asymptomatic SARS-CoV-2 infection. **Molecules**, v. 26, n. 5, p. 1200, 2021.

CIEPLINSKI, T. et al. We should at least be able to design molecules that dock well. **arXiv preprint arXiv:2006.16955**, 2020.

CIMINI, A. et al. Cocoa powder triggers neuroprotective and preventive effects in a human Alzheimer's disease model by modulating BDNF signaling pathway. **Journal of Cellular Biochemistry**, v. 114, n. 10, p. 2209–2220, out. 2013.

COOPER, K. A. et al. Cocoa and health: A decade of research. **British Journal of Nutrition**, v. 99, n. 1, p. 1-11, jan. 2008.

COQ-HUELVA, D.; TORRES-NAVARRETE, B.; BUENO-SUÁREZ, C. Indigenous worldviews and Western conventions: Sumak Kawsay and cocoa production in Ecuadorian Amazonia. **Agriculture and Human Values**, v. 35, n. 1, p. 163–179, 1 mar. 2018.

CORLEY, D. G.; DURLEY, R. C. Strategies for database dereplication of natural products. **Journal of natural products**, v. 57, n. 11, p. 1484–1490, 1994.

CORTEZ, D. et al. Changes in bioactive compounds during fermentation of cocoa (Theobroma cacao) harvested in Amazonas-Peru. **Current Research in Food Science**, v. 6, p.1000494, jan. 2023.

COSTANZO, M. J. et al. Potent, small-molecule inhibitors of human mast cell tryptase. Antiasthmatic action of a dipeptide-based transition-state analogue containing a benzothiazole ketone. **Journal of medicinal chemistry**, v. 46, n. 18, p. 3865–3876, 2003a.

COVA, T. F. G. G.; PAIS, A. A. C. C. Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. **Frontiers in Chemistry**, v. 7, p.809, nov. 2019.

CRAGG, G. M.; NEWMAN, D. J. Natural products: a continuing source of novel drug leads. **Biochimica et Biophysica Acta (BBA)-General Subjects**, v. 1830, n. 6, p. 3670–3695, 2013.

CROTEAU, R. et al. Natural products (secondary metabolites). **Biochemistry and molecular biology of plants**, v. 24, p. 1250–1319, 2000.

CURTIS, M. J. et al. Planning experiments: Updated guidance on experimental design and analysis and their reporting III. **British Journal of Pharmacology**, v. 179, n. 15, p. 3907-3913, ago. 2022.

DA SILVEIRA, N. J. F. et al. Web services for molecular docking simulations. Em: **Docking Screens for Drug Discovery**, p. 221-229, 2019.

DAS, S.; DEY, A.; ROY, N. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. **International Journal of Computer Applications**, v. 115, n. 9, 2015.

DASIMAN, R. et al. A Review of Procyanidin: Updates on current bioactivities and potential health benefits. **Biointerface Res. Appl. Chem**, v. 12, n. 5, p. 5918–5940, 2022.

DAYAN, P.; NIV, Y. Reinforcement learning: The Good, The Bad and The Ugly. **Current Opinion in Neurobiology**, v. 18, n. 2, p. 185-196, abr. 2008.

DE BRITO, E. S. et al. Structural and chemical changes in cocoa (Theobroma cacao L) during fermentation, drying and roasting. **Journal of the Science of Food and Agriculture**, v. 81, n. 2, p. 281–288, 2001.

DE QUEIROZ, L. N. et al. New substances of Equisetum hyemale L. extracts and their in vivo antitumoral effect against oral squamous cell carcinoma. **Journal of Ethnopharmacology**, v.303, p. 116043, 2022.

DE VUYST, L.; LEROY, F. Functional role of yeasts, lactic acid bacteria and acetic acid bacteria in cocoa fermentation processes. **FEMS Microbiology Reviews**, v. 44, n. 4, p. 432–453, 2020.

DE VUYST, L.; WECKX, S. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. **Journal of Applied Microbiology**, v. 121, n. 1, p. 5–17, 2016a.

DEL CAMPO, M.; CARLSON, A.; MANNINGER, S. Towards Hallucinating Machines - Designing with Computational Vision. **International Journal of Architectural Computing**, v. 19, n. 1, p. 88–103, mar. 2021.

DELGADO-OSPINA, J. et al. The role of fungi in the cocoa production chain and the challenge of climate change. **Journal of Fungi**, v. 7, n. 3, p. 202, 2021.

DEMAIN, A. L. Importance of microbial natural products and the need to revitalize their discovery. **Journal of Industrial Microbiology and Biotechnology**, v. 41, n. 2, p. 185–201, 2014.

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: **2009 IEEE conference on computer vision and pattern recognition**. Ieee, 2009, p. 248-255.

DEUS, V. L. et al. Understanding amino acids and bioactive amines changes during on-farm cocoa fermentation. **Journal of Food Composition and Analysis**, v. 97, p. 103776, abr. 2021.

DIFRANCISCO-DONOGHUE, J. et al. Effects of Tyrosine on Parkinson's Disease: A Randomized, Double-Blind, Placebo-Controlled Trial. **Movement Disorders Clinical Practice**, v. 1, n. 4, p. 348–353, dez. 2014.

DMITRYJUK, M.; ŁOPIEŃSKA-BIERNAT, E.; FARJAN, M. The level of sugars and synthesis of trehalose in Ascaris suum tissues. **Journal of Helminthology**, v. 83, n. 3, p. 237–243, 2009.

DONG, X. et al. Web service infrastructure for chemoinformatics. **Journal of chemical information and modeling**, v. 47, n. 4, p. 1303–1307, 2007.

DORFMAN, L. J.; JARVIK, M. E. Comparative stimulant and diuretic actions of caffeine and theobromine in man. **Clinical Pharmacology & Therapeutics**, v. 11, n. 6, p. 869–872, 1970.

DREW, K. L. M. et al. Size estimation of chemical space: how big is it? **Journal of Pharmacy and Pharmacology**, v. 64, n. 4, p. 490–495, 2012.

DRICHE, E. H. et al. A new Streptomyces strain isolated from Saharan soil produces di-(2-ethylhexyl) phthalate, a metabolite active against methicillin-resistant Staphylococcus aureus. **Annals of microbiology**, v. 65, n. 3, p. 1341–1350, 2015.

DU, X. et al. Insights into protein–ligand interactions: Mechanisms, models, and methods. **International Journal of Molecular Sciences**, v. 17, n. 2, p. 144, 2016.

DUCA, D. et al. Indole-3-acetic acid in plant–microbe interactions. **Antonie Van Leeuwenhoek**, v. 106, n. 1, p. 85–125, 2014.

DUCA, D. R.; GLICK, B. R. Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. **Applied microbiology and biotechnology**, v. 104, p. 8607–8619, 2020.

DÜHRKOP, K. et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. **Nature methods**, v. 16, n. 4, p. 299–302, 2019.

DUKE, J. A. Biological activity summary for cocoa (Theobroma cacao L.). **Journal of Medicinal Food**, v. 3, n. 2, p. 115–119, 2000.

DZOBO, K. The Role of Natural Products as Sources of Therapeutic Agents for Innovative Drug Discovery. **Comprehensive Pharmacology**, p. 408, 2022.

EALES, J. et al. Human health impacts of exposure to phthalate plasticizers: An overview of reviews. **Environment** International, v. 158, p. 106903, jan. 2022.

EKINS, S.; CLARK, A. M.; WILLIAMS, A. J. Open drug discovery teams: A chemistry mobile app for collaboration. **Molecular Informatics**, v. 31, n. 8, p. 585–597, ago. 2012.

ELHARROUSS, O. et al. Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches. **arXiv preprint arXiv:2206.08016**, jun. 2022.

ELLAM, S.; WILLIAMSON, G. Cocoa and human health. **Annual Review of Nutrition**, v. 33, p. 105–128, jul. 2013.

ERTL, P.; ROHDE, B.; SELZER, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. **Journal of medicinal chemistry**, v. 43, n. 20, p. 3714–3717, 2000.

FEBRIANTO, N. A.; ZHU, F. Composition of methylxanthines, polyphenols, key odorant volatiles and minerals in 22 cocoa beans obtained from different geographic origins. **LWT**, v. 153, p. 112395, jan. 2022.

FEDURAEV, P. et al. Phenylalanine and tyrosine as exogenous precursors of wheat (triticum aestivum L.) secondary metabolism through PAL-associated pathways. **Plants**, v. 9, n. 4, p. 476, abr. 2020.

FELDMANN, C. et al. Identifying promiscuous compounds with activity against different target classes. **Molecules**, v. 24, n. 22, p. 4185, 2019.

FERNSTROM, J. D.; FERNSTROM, M. H. Tyrosine, Phenylalanine, and Catecholamine Synthesis and Function in the Brain. **The Journal of Nutrition**, v. 137, n. 6, p. 1539S-1547S, 2007.

FIGUEROA-HERNÁNDEZ, C. et al. The challenges and perspectives of the selection of starter cultures for fermented cocoa beans. **International Journal of Food Microbiology**, v. 301, p. 41–50, 2019.

FRANÇOIS, J.; PARROU, J. L. Reserve carbohydrates metabolism in the yeast Saccharomyces cerevisiae. **Fems microbiology reviews**, v. 25, n. 1, p. 125–145, 2001.

FRANZEN, M.; BORGERHOFF MULDER, M. Ecological, economic and social perspectives on cocoa production worldwide. **Biodiversity and Conservation**, v. 16, p. 3835-3849, dez. 2007.

FREDHOLM, B. B.; SMIT, H. J. Theobromine and the pharmacology of cocoa. **Methylxanthines**, p. 201–234, 2011.

FREIESLEBEN, J.; KEIM, J.; GRUTSCH, M. Machine learning and Design of Experiments: Alternative approaches or complementary methodologies for quality improvement? **Quality and Reliability Engineering International**, v. 36, n. 6, p. 1837–1848, out. 2020.

FURUSHIMA, D. et al. Prevention of acute upper respiratory infections by consumption of catechins in healthcare workers: A randomized, placebo-controlled trial. **Nutrients**, v. 12, n. 1, p. 4, jan. 2019.

FUSAR-POLI, L. et al. The effect of cocoa-rich products on depression, anxiety, and mood: A systematic review and meta-analysis. **Critical Reviews in Food Science and Nutrition**, v. 62, n. 28, p. 7905-7916, 2022.

GALLEGO, A. M. et al. Transcriptomic analyses of cacao flavonoids produced in photobioreactors. **BMC genomics**, v. 22, n. 1, p. 1–18, 2021.

GARCÍA-ORTEGÓN, M. et al. DOCKSTRING: easy molecular docking yields better benchmarks for ligand design. **Journal of Chemical Information and Modeling**, v. 62, n. 15, p. 3486-3502, ago. 2022.

GARG, A. K. et al. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. **Proceedings of the National Academy of Sciences**, v. 99, n. 25, p. 15898–15903, 2002.

GARIBOTTO, G. et al. The metabolic conversion of phenylalanine into tyrosine in the human kidney: Does it have nutritional implications in renal patients? **Journal of Renal Nutrition**, v. 12, n. 1, p. 8–16, jan. 2002.

GAUDÊNCIO, S. P. et al. Advanced Methods for Natural Products Discovery: Bioactivity Screening, Dereplication, Metabolomics Profiling, Genomic Sequencing,

Databases and Informatic Tools, and Structure Elucidation. **Marine Drug**, v.21, n. 5, p. 308, 2023.

GAWEHN, E. et al. Advancing drug discovery via GPU-based deep learning. **Expert Opinion on Drug Discovery**, v. 13, n. 7, p. 579-582, jul. 2018.

GENHEDEN, S.; RYDE, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. **Expert Opinion on Drug Discovery** v. 10, n. 5, p. 449-461, 2015.

GHAYUR, M. N.; KHAN, H.; GILANI, A. H. Antispasmodic, bronchodilator and vasodilator activities of (+)-catechin, a naturally occurring flavonoid. **Archives of pharmacal research**, v. 30, p. 970-975, 2007.

GHIANDONI, G. M.; CALDEWEYHER, E. Fast calculation of hydrogen-bond strengths and free energy of hydration of small molecules. **Scientific Reports**, v. 13, n. 1, dez. 2023.

GHOSH, D. **A cinnamon-derived procyanidin type-A compound: A potential candidate molecule against coronaviruses including COVID-19**. **Journal of Ayurveda Case Reports**, v. 3, n. 4, p. 122-126, 2020.

GOETZ, M. et al. Extremely randomized trees based brain tumor segmentation. **Proceeding of BRATS challenge-MICCAI**, v. 14, p. 6–11, 2014.

GOHLKE, H.; KLEBE, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. **Angewandte Chemie International Edition**, v. 41, n. 15, p. 2644–2676, 2002.

GONZÁLEZ, M. P. et al. BCUT descriptors to predicting affinity toward A3 adenosine receptors. **Bioorganic & medicinal chemistry letters**, v. 15, n. 15, p. 3491–3495, 2005.

GORMAN, J. W.; HINMAN, J. E. Simplex lattice designs for multicomponent systems. **Technometrics**, v. 4, n. 4, p. 463–487, 1962.

GRASSI, D. et al. Cocoa reduces blood pressure and insulin resistance and improves endothelium-dependent vasodilation in hypertensives. **Hypertension**, v. 46, n. 2, p. 398–405, ago. 2005.

GROMSKI, P. S. et al. How to explore chemical space using algorithms and automation. **Nature Reviews Chemistry**, v. 3, n. 2, p. 119–128, 2019.

GU, J. et al. Recent advances in convolutional neural networks. **Pattern recognition**, v. 77, p. 354–377, 2018.

GUÀRDIA, E. et al. A molecular dynamics simulation study of hydrogen bonding in aqueous ionic solutions. **Journal of Molecular Liquids**, v. 117, n. 1-3, p. 63-67, 2005.

GUEHI, T. S. et al. Performance of different drying methods and their effects on the chemical quality attributes of raw cocoa material. **International Journal of Food Science and Technology**, v. 45, n. 8, p. 1564–1571, ago. 2010.

GUPTA, R. et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. **Molecular Diversity**, v. 25, n. 3, p. 1315–1360, 2021.

GUTIÉRREZ-RÍOS, H. G. et al. Yeasts as Producers of Flavor Precursors during Cocoa Bean Fermentation and Their Relevance as Starter Cultures: A Review. **Fermentation**, v. 8, n. 7, p. 331, 2022.

HABIB, M. R.; KARIM, M. R.; OTHERS. Antitumour evaluation of di-(2-ethylhexyl) phthalate (DEHP) isolated from Calotropis gigantea L. flower. **Acta Pharm**, v. 62, n. 4, p. 607–615, 2012.

HALL, T. et al. Structure of human ADAM-8 catalytic domain complexed with batimastat. **Acta Crystallographica Section F: Structural Biology and Crystallization Communications**, v. 68, n. 6, p. 616–621, 2012a.

HASTINGS, J. et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. **Nucleic acids research**, v. 41, n. D1, p. D456–D463, 2012.

HAUG, K. et al. MetaboLights: a resource evolving in response to the needs of its scientific community. **Nucleic acids research**, v. 48, n. D1, p. D440–D444, 2020.

HE, X. et al. Trehalose Alleviates Crystalline Silica-Induced Pulmonary Fibrosis via Activation of the TFEB-Mediated Autophagy-Lysosomal System in Alveolar Macrophages. **Cells**, v. 9, n. 1, jan. 2020.

HEINRICH, M. Ethnobotany and natural products: the search for new molecules, new treatments of old diseases or a better understanding of indigenous cultures? **Current Topics in Medicinal Chemistry**, v. 3, n. 2, p. 141–154, 2003.

HELGUERA, A. M. et al. Applications of 2D descriptors in drug design: a DRAGON tale. **Current topics in medicinal chemistry**, v. 8, n. 18, p. 1628–1655, 2008.

HENSS, L. et al. The green tea catechin epigallocatechin gallate inhibits SARS-CoV-2 infection. **The Journal of general virology**, v. 102, n. 4, 2021.

HESHMATISAFA, S.; SEPPÄNEN, M. Exploring API-driven business models: Lessons learned from Amadeus's digital transformation. **Digital Business**, v. 3, n. 1, p. 100055, jun. 2023.

HILBE, J. M. STATISTICA 7: an overview. **The American Statistician**, v. 61, n. 1, p. 91–94, 2007.

HILLMAN, E. T.; READNOUR, L. R.; SOLOMON, K. V. **Exploiting the natural product potential of fungi with integrated-omics and synthetic biology approaches**. **Current Opinion in Systems Biology**, v. 5, p. 50-56, 2017.

HOANG, V. L. T.; LI, Y.; KIM, S.-K. Cathepsin B inhibitory activities of phthalates isolated from a marine Pseudomonas strain. **Bioorganic & medicinal chemistry letters**, v. 18, n. 6, p. 2083–2088, 2008.

HOLLINGSWORTH, S. A.; DROR, R. O. **Molecular Dynamics Simulation for All**. **Neuron**, v. 99, n. 6, p. 1129-1143, 2018.

HOLTEN-ANDERSEN, L. et al. Combination of the cationic surfactant dimethyl dioctadecyl ammonium bromide and synthetic mycobacterial cord factor as an efficient adjuvant for tuberculosis subunit vaccines. **Infection and immunity**, v. 72, n. 3, p. 1608–1617, 2004.

HORWOOD, J.; NOUTAHI, E. Molecular Design in Synthetically Accessible Chemical Space via Deep Reinforcement Learning. **ACS Omega**, v. 5, n. 51, p. 32984–32994, dez. 2020.

HOU, T. et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. **Journal of Chemical Information and Modeling**, v. 51, n. 1, p. 69–82, jan. 2011.

HU, Y.; BAJORATH, J. Compound promiscuity: What can we learn from current data? **Drug Discovery Today**, v. 18, n. 13-14, p. 644-650, 2013.

HUANG, B.; ZHANG, Y. Teaching an old dog new tricks: Drug discovery by repositioning natural products and their derivatives. **Drug Discovery Today**, v. 27, n. 7, p. 1936-1944, 2022.

HUANG, L. et al. Phthalic acid esters: Natural sources and biological activities. **Toxins**, v. 13, n. 7, p. 495, 2021.

HUANG, M.; LU, J.-J.; DING, J. Natural products in cancer therapy: past, present and future. **Natural Products and Bioprospecting**, v. 11, n. 1, p. 5–13, 2021.

HUBBARD, R. E.; KAMRAN HAIDER, M. Hydrogen Bonds in Proteins: Role and Strength. Em: **eLS**. [s.l.] Wiley, 2010.

HUBER, F. et al. matchms-processing and similarity evaluation of mass spectrometry data. **bioRxiv**, p. 2020.08. 06.239244, 2020.

HUEY, R.; MORRIS, G. M.; FORLI, S. Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. **The Scripps Research Institute Molecular Graphics Laboratory**, v. 10550, p. 92037, 2012.

HUNTER, S. V. Analysing and representing narrative data: The long and winding road. **Current narratives**, v. 1, n. 2, p. 44–54, 2010.

IGAWA, T. K.; DE TOLEDO, P. M.; ANJOS, L. J. S. Climate change could reduce and spatially reconfigure cocoa cultivation in the Brazilian Amazon by 2050. **PLoS ONE**, v. 17, n. 1 January, jan. 2022.

ITURRIAGA, G.; SUÁREZ, R.; NOVA-FRANCO, B. Trehalose metabolism: From osmoprotection to signaling. **International Journal of Molecular Sciences**, v. 10, n. 9, p. 3793-3810, 2009.

JAIN, R. S. et al. Review on methylxanthine, theobromine and theophylline. **Asian Journal of Pharmaceutical Analysis**, v. 10, n. 3, p. 173–174, 2020.

JIANG, Y. et al. Procyanidin B2 Suppresses Lipopolysaccharides-Induced Inflammation and Apoptosis in Human Type II Alveolar Epithelial Cells and Lung Fibroblasts. **Journal of Interferon and Cytokine Research**, v. 40, n. 1, p. 54–63, 1 jan. 2020.

JOHN, W. A. et al. Experimentally modelling cocoa bean fermentation reveals key factors and their influences. **Food Chemistry**, v. 302, n. July 2019, p. 125335, 2020.

JORGENSEN, W. L. The Many Roles of Computation in Drug Discovery. **Science**, v. 303, n. 5665, p. 1813-1818, 2004.

JOUANEH, T. M. M. et al. Incorporating LC-MS/MS Analysis and the Dereplication of Natural Product Samples into an Upper-Division Undergraduate Laboratory Course. **Journal of Chemical Education**, v. 99, n. 7, p. 2636–2642, jul. 2022.

KADOW, D. et al. Fermentation-like incubation of cocoa seeds (Theobroma cacao L.) - Reconstruction and guidance of the fermentation process. **LWT**, v. 62, n. 1, p. 357–361, 1 jun. 2015.

KANG, Y. et al. Natural language processing (NLP) in management research: A literature review. **Journal of Management Analytics**, v. 7, n. 2, p. 139-172, 2020.

KANWAL et al. Indole-3-acetamides: As Potential Antihyperglycemic and Antioxidant Agents; Synthesis, in Vitro α-Amylase Inhibitory Activity, Structure-Activity Relationship, and in Silico Studies. **ACS Omega**, v. 6, n. 3, p. 2264–2275, jan. 2021.

KATZ, D. L.; DOUGHTY, K.; ALI, A. Cocoa and chocolate in human health and disease. **Antioxidants and Redox Signaling**, v. 15, n. 10, p. 2779-2811, 2011.

KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. **Advances in neural information processing systems**, v. 30, 2017.

KEITH, J. A. et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. **Chemical reviews**, v. 121, n. 16, p. 9816–9872, 2021.

KHAN, N. et al. Cocoa polyphenols and inflammatory markers of cardiovascular disease. **Nutrients**, v. 6, n. 2, p. 844–880, 2014.

KHANFAR, M. A.; TAHA, M. O. Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR inhibitors. **Journal of chemical information and modeling**, v. 53, n. 10, p. 2587–2612, 2013.

KIEFER, B. A. et al. The identification of adenine in cacao products. **Journal of Liquid Chromatography**, v. 6, n. 5, p. 927–930, abr. 1983.

KIND, T.; FIEHN, O. Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. **Phytochemistry Letters**, v. 21, p. 313-319, set. 2017.

KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature reviews Drug discovery**, v. 3, n. 11, p. 935–949, 2004.

KJELDAHL, K.; BRO, R. Some common misunderstandings in chemometrics. **Journal of Chemometrics**, v. 24, n. 7–8, p. 558–564, 2010.

KOEHN, F. E.; CARTER, G. T. The evolving role of natural products in drug discovery. **Nature reviews Drug discovery**, v. 4, n. 3, p. 206–220, 2005.

KONGOR, J. E. et al. Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile - A review. **Food Research International**, v. 82, p. 42-52, 2016.

KOULOURIDI, E. et al. A primer on natural product-based virtual screening. **Physical Sciences Reviews**, v. 4, n. 6, 2019.

KOYAMA, Y. et al. Metabolism of purine bases, nucleosides and alkaloids in theobromine-forming Theobroma cacao leaves. **Plant Physiology and Biochemistry**, v. 41, n. 11–12, p. 977–984, 2003.

KRZYWINSKI, M.; ALTMAN, N. Classification and regression trees. **Nature Methods**, v. 14, n. 8, p. 757–758, 2017.

KULKARNI, G. B. et al. Indole-3-acetic acid biosynthesis in Fusarium delphinoides strain GPK, a causal agent of Wilt in Chickpea. **Applied biochemistry and biotechnology**, v. 169, n. 4, p. 1292–1305, 2013.

KULKARNI VISHAKHA, S.; BUTTE KISHOR, D.; RATHOD SUDHA, S. Natural polymers–A comprehensive review. **International journal of research in pharmaceutical and biomedical sciences**, v. 3, n. 4, p. 1597–1613, 2012.

KUMAR, S. et al. Discovery of New Hydroxyethylamine Analogs against 3CLproProtein Target of SARS-CoV-2: Molecular Docking, Molecular Dynamics Simulation, and Structure-Activity Relationship Studies. **Journal of Chemical Information and Modeling**, v. 60, n. 12, p. 5754–5770, dez. 2020.

KUMARI, R.; KUMAR, R.; LYNN, A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1951–1962, jul. 2014.

KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions. **Journal of molecular biology**, v. 161, n. 2, p. 269–288, 1982.

LABUTE, P. A widely applicable set of descriptors. **Journal of Molecular Graphics and Modelling**, v. 18, n. 4–5, p. 464–477, 2000.

LAMBRAKIS, D. P. Experiments with mixtures: A generalization of the simplex-lattice design. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 30, n. 1, p. 123–136, 1968.

LAMBROT, R. et al. Phthalates impair germ cell development in the human fetal testis in vitro without change in testosterone production. **Environmental health perspectives**, v. 117, n. 1, p. 32–37, 2009.

LANDRUM, G.; OTHERS. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. **Greg Landrum**, v. 8, p. 31, 2013.

LANGLYKKE, A. CRC Handbook of antibiotic compound (IV), Edited by Bardy J et al. **CRC, Boca Raton, FL**, 1980.

LEACH, A. R.; SHOICHET, B. K.; PEISHOFF, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. **Journal of Medicinal Chemistry**, v. 49, n. 20, p. 5851-5855, 2006..

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.

LEE, H. W.; CHOI, I. W.; HA, S. K. Immunostimulatory Activities of Theobromine on Macrophages via the Activation of MAPK and NF-κB Signaling Pathways. **Current Issues in Molecular Biology**, v. 44, n. 9, p. 4216–4228, set. 2022.

LI, C. et al. Blocking the 4-1BB pathway ameliorates crystalline silica-induced lung inflammation and fibrosis in mice. **Theranostics**, v. 6, n. 12, p. 2052–2067, 2016.

LI MANNI, G. et al. The OpenMolcas Web: A Community-Driven Approach to Advancing Computational Chemistry. **Journal of Chemical Theory and Computation**, 2023.

LIMA, L. J. R. et al. Theobroma cacao L.,"the food of the gods": quality determinants of commercial cocoa beans, with particular reference to the impact of fermentation. **Critical reviews in food science and nutrition**, v. 51, n. 8, p. 731–761, 2011.

LIN, L.; XU, X. Indole-3-acetic acid production by endophytic Streptomyces sp. En-1 isolated from medicinal plants. **Current Microbiology**, v. 67, n. 2, p. 209–217, 2013.

LIU, Z. et al. DeepScreening: A deep learning-based screening web server for accelerating drug discovery. **Database**, v. 2019, n. 1, 2019.

LOTFY, W. A. et al. Production of di-(2-ethylhexyl) phthalate by Bacillus subtilis AD35: Isolation, purification, characterization and biological activities. **Microbial Pathogenesis**, v. 124, p. 89–100, nov. 2018.

LOWE, D. Chemical space is big. Really big. **MedChemComm**, v. 6, n. 1, p. 12, 2015.

LUNN, J. E. et al. Trehalose metabolism in plants. **The Plant Journal**, v. 79, n. 4, p. 544–567, 2014.

MACHONIS, P. R. et al. Method for the determination of catechin and epicatechin enantiomers in cocoa-Based ingredients and products by High-Performance Liquid chromatography: Single- Laboratory validation. **Journal of AOAC International**, v. 95, n. 2, p. 500–507, mar. 2012.

MAJUMDER, R.; MANDAL, M. Screening of plant-based natural compounds as a potential COVID-19 main protease inhibitor: an in silico docking and molecular

dynamics simulation approach. **Journal of Biomolecular Structure and Dynamics**, v. 40, n. 2, p. 696–711, 2022.

MALLMANN, L. P.; DE OLIVEIRA RIOS, A.; RODRIGUES, E. MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data. **Food Research International**, v. 163, p. 112315, 2023.

MARKOVIC, S.; GUTMAN, I. Spectral moments of the edge adjacency matrix in molecular graphs. Benzenoid hydrocarbons. **J. Chem. Inf. Comput. Sci.**, v. 39, n. 2, p. 289–293, 1999.

MARTÍNEZ-PINILLA, E.; OÑATIBIA-ASTIBIA, A.; FRANCO, R. The relevance of theobromine for the beneficial effects of cocoa consumption. **Frontiers in Pharmacology**, v. 6, p. 30, 2015.

MARTINON, D. et al. Potential fast COVID-19 containment with trehalose. **Frontiers in Immunology**, v. 11, p. 1623, 2020.

MATISSEK, R. Evaluation of xanthine derivatives in chocolate–nutritional and chemical aspects. **Zeitschrift für Lebensmitteluntersuchung und-Forschung A**, v. 205, p. 175–184, 1997.

MAUN, H. R. et al. Bivalent antibody pliers inhibit β-tryptase by an allosteric mechanism dependent on the IgG hinge. **Nature communications**, v. 11, n. 1, p. 1–12, 2020.

MCLAUGHLIN, C. A. et al. Regression of line-10 hepatocellular carcinomas following treatment with water-soluble, microbial extracts combined with trehalose or arabinose mycolates. **Cancer Immunology, Immunotherapy**, v. 4, n. 1, p. 61–68, 1978.

MEDEMA, M. H.; FISCHBACH, M. A. Computational approaches to natural product discovery. Nature Chemical Biology. **Nature**, v. 11, n. 9, p. 639-648, 2015.

MEUWLY, M. Machine learning for chemical reactions. **Chemical Reviews**, v. 121, n. 16, p. 10218-10239, 2021.

MIN, S.; LEE, B.; YOON, S. Deep learning in bioinformatics. **Briefings in bioinformatics**, v. 18, n. 5, p. 851-869, 2017.

MISHRA, C. B. et al. Identifying the natural polyphenol catechin as a multi-targeted agent against SARS-CoV-2 for the plausible therapy of COVID-19: an integrated computational approach. **Briefings in Bioinformatics**, v. 22, n. 2, p. 1346–1360, 2021.

MOHIMANI, H. et al. Dereplication of peptidic natural products through database search of mass spectra. **Nature chemical biology**, v. 13, n. 1, p. 30–37, 2017.

MONTGOMERY, D. C. The Use of Statistical Process Control and Design of Experiments in Product and Process Improvement. **IIE Transactions (Institute of Industrial Engineers)**, v. 24, n. 5, p. 4–17, 1992.

MOREIRA, I. M. DA V. et al. Microbial succession and the dynamics of metabolites and sugars during the fermentation of three different cocoa (Theobroma cacao L.) hybrids. **Food Research International**, v. 54, n. 1, p. 9–17, nov. 2013.

MORENO-ZAMBRANO, M. et al. A mathematical model of cocoa bean fermentation. **Royal Society Open Science**, v. 5, n. 10, 1 out. 2018.

MORIWAKI, H. et al. Mordred: a molecular descriptor calculator. **Journal of cheminformatics**, v. 10, n. 1, p. 1–14, 2018.

MORRONE XAVIER, M. et al. SAnDReS a computational tool for statistical analysis of docking results and development of scoring functions. **Combinatorial chemistry & high throughput screening**, v. 19, n. 10, p. 801–812, 2016.

MOTAMAYOR, J. C. et al. Cacao domestication I: The origin of the cacao cultivated by the Mayas. **Heredity**, v. 89, n. 5, p. 380–386, nov. 2002.

MOUSHUMI PRIYA, A.; JAYACHANDRAN, S. Induction of apoptosis and cell cycle arrest by Bis (2-ethylhexyl) phthalate produced by marine Bacillus pumilus MB 40. **Chemico-Biological Interactions**, v. 195, n. 2, p. 133–143, jan. 2012.

MOZZI, FERNANDA.; RAYA, R. R.; VIGNOLO, G. M. **Biotechnology of lactic acid bacteria. 2**. Singapore: Wiley-Blackwell, 2015.

MYERS, R. H. et al. Response Surface Methodology: A Retrospective and Literature Survey. **Journal of Quality Technology**, v. 36, n. 1, p. 53-77, 2004.

NABAVI, S. et al. Anti-Oxidative Polyphenolic Compounds of Cocoa. **Current Pharmaceutical Biotechnology**, v. 16, n. 10, p. 891–901, 2015.

NEHLIG, A. The neuroprotective effects of cocoa flavanol and its influence on cognitive performance. **British journal of clinical pharmacology**, v. 75, n. 3, p. 716–727, 2013.

NET, S. et al. Occurrence, fate, behavior and ecotoxicological state of phthalates in different environmental matrices. **Environmental Science and Technology**, v. 49, n. 7, p. 4019-4035, 2015.

NETTLES, J. H. et al. Bridging chemical and biological space:“target fishing” using 2D and 3D molecular descriptors. **Journal of medicinal chemistry**, v. 49, n. 23, p. 6802–6810, 2006.

NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. **Journal of natural products**, v. 75, n. 3, p. 311–335, 2012.

NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs from 1981 to 2014. **Journal of natural products**, v. 79, n. 3, p. 629–661, 2016.

NOBLE, W. S. What is a support vector machine? **Nature biotechnology**, v. 24, n. 12, p. 1565–1567, 2006.

NOLIBE, D. et al. Activation of rat alveolar macrophages and protection against iv injected tumor cells by intratracheal administration of trehalose dimycolate. **Cancer Immunology, Immunotherapy**, v. 23, n. 3, p. 200–206, 1986.

NOSRATI, M. Python: An appropriate language for real world programming. **World Applied Programming**, v. 1, n. 2, p. 110–117, 2011.

NOVAK, J. et al. Can natural products stop the SARS-CoV-2 virus? A docking and molecular dynamics study of a natural product database. **Future Medicinal Chemistry**, v. 13, n. 4, p. 363–378, 1 fev. 2021.

nsb0902-646. [s.d.].

OHTAKE, S.; WANG, Y. J. Trehalose: Current use and future applications. **Journal of Pharmaceutical Sciences**, v. 100, n. 6, p. 2020-2053, 2011.

ONG, S. P. et al. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. **Computational Materials Science**, v. 97, p. 209–215, fev. 2015.

ONUFRIEV, A. V; CASE, D. A. Generalized Born Implicit Solvent Models for Biomolecules. **Annual review of biophysics**, v. 48, p. 275-296, 2019.

ORTIZ, A.; SANSINENEA, E. Di-2-ethylhexylphthalate may be a natural product, rather than a pollutant. **Journal of Chemistry**, v. 2018, 2018.

OUATTARA, H. G. et al. Molecular identification and pectate lyase production by Bacillus strains involved in cocoa fermentation. **Food Microbiology**, v. 28, n. 1, p. 1–8, fev. 2011.

OZATO, N. et al. Effect of catechins on upper respiratory tract infections in winter: A randomized, placebo-controlled, double-blinded trial. **Nutrients**, v. 14, n. 9, p. 1856, 2022.

ÖZDEMIR, V.; HEKIM, N. Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, “the Internet of Things” and Next-Generation Technology Policy. **OMICS A Journal of Integrative Biology**, v. 22, n. 1, p. 65–76, 2018.

PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. **Biophysical reviews**, v. 9, n. 2, p. 91–102, 2017.

PAJOKH, M.; POURFRIDONI, M. Proposing a nasal trehalose-induced autophagy approach against SARS-CoV 2. **Health Science Reports**, v. 4, n. 3, 2021.

PAN, G. et al. Decreased serum free testosterone in workers exposed to high levels of di-n-butyl phthalate (DBP) and di-2-ethylhexyl phthalate (DEHP): a cross-sectional study in China. **Environmental health perspectives**, v. 114, n. 11, p. 1643–1648, 2006.

PAPALEXANDRATOU, Z. et al. Hanseniaspora opuntiae, Saccharomyces cerevisiae, Lactobacillus fermentum, and Acetobacter pasteurianus predominate during well-performed Malaysian cocoa bean box fermentations, underlining the importance of these microbial species for a successful cocoa bean fermentation process. **Food Microbiology**, v. 35, n. 2, p. 73–85, set. 2013.

PARANT, M. et al. Enhancement of Nonspecific Immunity to Bacterial Infection by Cord Factor (6, 6′ MTrehalose Dimycolate). **Journal of Infectious Diseases**, v. 135, n. 5, p. 771–777, 1977.

PASSOS, F. M. L.; LOPEZ, A. S.; SILVA, D. O. Aeration and its influence on the microbial sequence in cacao fermentations in Bahia, with emphasis on lactic acid bacteria. **Journal of Food Science**, v. 49, n. 6, p. 1470–1474, 1984.

PAUL, D. et al. Artificial intelligence in drug discovery and development. **Drug Discovery Today**, v. 26, n. 1, p. 80, 2021.

PIEPEL, G. F.; CORNELL, J. A. Designs for Mixture-Amount Experiments. **Journal of Quality Technology**, v. 19, n. 1, p. 11–28, jan. 1987.

PIEPEL, G. F.; CORNELL, J. A. Mixture experiment approaches: examples, discussion, and recommendations. **Journal of Quality Technology**, v. 26, n. 3, p. 177–196, 1994.

PILON, A. C. et al. NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. **Scientific Reports**, v. 7, n. 1, p. 1–12, 2017.

POLITIS, S. N. et al. Design of experiments (DoE) in pharmaceutical development. **Drug Development and Industrial Pharmacy**, v. 43, n. 6, p. 889-901, 2017.

PUERARI, C.; MAGALHÃES, K. T.; SCHWAN, R. F. New cocoa pulp-based kefir beverages: Microbiological, chemical composition and sensory analysis. **Food Research International**, v. 48, n. 2, p. 634–640, out. 2012.

QIAN, Z.-J.; KANG, K.-H.; KIM, S.-K. Isolation and antioxidant activity evaluation of two new phthalate derivatives from seahorse, Hippocampus kuda Bleeler. **Biotechnology and Bioprocess Engineering**, v. 17, n. 5, p. 1031–1040, 2012.

QIN, G. F. et al. MS/MS-Based Molecular Networking: An Efficient Approach for Natural Products Dereplication. **Molecules**, v. 28, n. 1, p. 157, 2022.

QIU, J. et al. A survey of machine learning for big data processing. **Eurasip Journal on Advances in Signal Processing**, v. 2016, p. 1-16, 2016.

RAHARDJO, Y. P. et al. Impact of controlled fermentation on the volatile aroma of roasted cocoa. **Brazilian Journal of Food Technology**, v. 25, 2022.

RAJASEKARAN, S.; RAJASEKAR, N.; SIVANANTHAM, A. Therapeutic potential of plant-derived tannins in non-malignant respiratory diseases. **The Journal of Nutritional Biochemistry**, v. 94, p. 108632, 2021.

RICHARDS, A. B. et al. Trehalose: a review of properties, history of use and human tolerance, and results of multiple safety studies. **Food and Chemical Toxicology**, v. 40, n. 7, p. 871-898, 2002.

RIED, K. et al. Effect of cocoa on blood pressure. **Cochrane Database of Systematic Reviews**, n. 8, 2012.

ROMANO, J. D.; TATONETTI, N. P. Informatics and computational methods in natural product drug discovery: A review and perspectives. **Frontiers in Genetics**, v. 10, p. 368, 2019..

ROSSETTI, G. G. et al. Non-covalent SARS-CoV-2 Mpro inhibitors developed from in silico screen hits. **Scientific reports**, v. 12, n. 1, p. 1–9, 2022.

ROTTIERS, H. et al. Dynamics of volatile compounds and flavor precursors during spontaneous fermentation of fine flavor Trinitario cocoa beans. **European Food Research and Technology**, v. 245, p. 1917–1937, 2019.

ROY, R. N. Bioactive natural derivatives of phthalate ester. **Critical reviews in biotechnology**, v. 40, n. 7, p. 913–929, 2020a.

ROY, R. N.; SEN, S. K. Fermentation studies for the production of dibutyl phthalate, an ester bioactive compound from Streptomyces albidoflavus MTCC 3662 using low-priced substrates. **Jordan J Biol Sci**, v. 6, p. 177–181, 2013.

RUDIN, C.; WAGSTAFF, K. L. Machine learning for science and society. Machine Learning. **Machine Learning**, v. 95, p. 1-9, 2014.

RUTZ, A. et al. The LOTUS initiative for open knowledge management in natural products research. **Elife**, v. 11, p. e70780, 2022.

SAITA, N. et al. Trehalose 6, 6′-dimycolate (cord factor) of Mycobacterium tuberculosis induces corneal angiogenesis in rats. **Infection and immunity**, v. 68, n. 10, p. 5991–5997, 2000.

SALDIVAR-GONZALEZ, F. I. et al. Chemical space and diversity of the NuBBE database: a chemoinformatic characterization. **Journal of Chemical Information and Modeling**, v. 59, n. 1, p. 74–85, 2018.

SALMASO, L. et al. Design of experiments and machine learning to improve robustness of predictive maintenance with application to a real case study. **Communications in Statistics: Simulation and Computation**, v. 51, n. 2, p. 570–582, 2022.

SANTANDER MUÑOZ, M. et al. An overview of the physical and biochemical transformation of cocoa seeds to beans and to chocolate: Flavor formation. **Critical Reviews in Food Science and Nutrition,** v. 60, n. 10, p. 1593-1613, 2020.

SARBU, I.; CSUTAK, O. The microbiology of cocoa fermentation. In: **Caffeinated and cocoa based beverages**. Woodhead Publishing, 2019. p. 423-446.

SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, v. 2, n. 3, p. 160, 2021.

SARUMATHI, S. et al. Statistica software: a state of the art review. **International Journal of Computer and Information Engineering**, v. 9, n. 2, p. 473–480, 2015.

SCHIRALDI, C.; DI LERNIA, I.; DE ROSA, M. Trehalose production: exploiting novel approaches. **TRENDS in Biotechnology**, v. 20, n. 10, p. 420–425, 2002.

SCHMIDT, M.; LIPSON, H. Distilling free-form natural laws from experimental data. **Science**, v. 324, n. 5923, p. 81–85, abr. 2009.

SCHROTH, G.; HARVEY, C. A. Biodiversity conservation in cocoa production landscapes: An overview. **Biodiversity and Conservation**, v. 16, p. 2237-2244, 2007.

SCHWAN, R. F. Cocoa fermentations conducted with a defined microbial cocktail inoculum. **Applied and Environmental Microbiology**, v. 64, n. 4, p. 1477–1483, 1998.

SCHWAN, R. F.; WHEALS, A. E. The microbiology of cocoa fermentation and its role in chocolate quality. **Critical Reviews in Food Science and Nutrition**, v. 44, n. 4, p. 205–221, 2004.

SHI, Y.-F. et al. Machine Learning for Chemistry: Basics and Applications. **Engineering**, jul. 2023.

SIMONS, F. E. R. et al. The bronchodilator effect and pharmacokinetics of theobromine in young patients with asthma. **Journal of allergy and clinical immunology**, v. 76, n. 5, p. 703–707, 1985.

SINGH, P.; RAJ, P.; NAMBOODIRI, V. P. EDS pooling layer. **Image and Vision Computing**, v. 98, p. 103923, jun. 2020.

ŚLEDŹ, P.; CAFLISCH, A. Protein structure-based drug design: from docking to molecular dynamics. **Current Opinion in Structural Biology**, v. 48, p. 93–102, 2018.

SONG, X. et al. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. **International Journal of Medical Informatics**, v. 151, p. 104484, 2021.

SOORI, M.; AREZOO, B.; DASTRES, R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. **Cognitive Robotics** jan. 2023.

SOROKINA, M. et al. COCONUT online: Collection of Open Natural Products database. **Journal of Cheminformatics**, v. 13, n. 1, p. 1-13, 2021.

SPARKS, T. C. et al. The new age of insecticide discovery-the crop protection industry and the impact of natural products. **Pesticide biochemistry and physiology**, v. 161, p. 12–22, 2019.

SQUEO, G. et al. Background, applications and issues of the experimental designs for mixture in the food sector. **Foods**, v. 10, n. 5, p. 1128, 2021.

SRINATH, K. R. Python–the fastest growing programming language. **International Research Journal of Engineering and Technology**, v. 4, n. 12, p. 354–357, 2017.

STANTON, D. T.; JURS, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. **Analytical Chemistry**, v. 62, n. 21, p. 2323–2329, 1990.

STENTA, M. **Chemistry 4.0: How the digital revolution is changing chemical research**. **Chimia**, v. 75, n. 3, p. 211-211, 2021.

SUD, M. et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. **Nucleic acids research**, v. 44, n. D1, p. D463–D470, 2016.

SUN, R. Y. Optimization for Deep Learning: An Overview. **Journal of the Operations Research Society of China**, v. 8, n. 2, p. 249–294, 1 jun. 2020.

SUSSMAN, J. L. et al. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. **Acta Crystallographica Section D: Biological Crystallography**, v. 54, n. 6, p. 1078–1084, 1998.

SUZUKI, T.; ASHIHARA, H.; WALLER, G. R. Purine and purine alkaloid metabolism in camellia and coffea plants**. Phytochemistry**, v. 31, n. 8, p. 2575-2584, 1992.

SVATIKOVA, A. et al. Circulating free nitrotyrosine in obstructive sleep apnea. **American Journal of Physiology-Regulatory, Integrative and Comparative Physiology**, v. 287, n. 2, p. R284-R287, 2004.

SVETNIK, V. et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. **Journal of chemical information and computer sciences**, v. 43, n. 6, p. 1947–1958, 2003.

SWAINSTON, N. et al. LibChEBI: An API for accessing the ChEBI database. **Journal of Cheminformatics**, v. 8, n. 1, mar. 2016.

TAMMINA, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. **International Journal of Scientific and Research Publications (IJSRP)**, v. 9, n. 10, p. 143–150, 2019.

TARKA, S. M.; CORNISH, H. H. The toxicology of cocoa and methylxanthines: a review of the literature. **CRC critical Reviews in Toxicology**, v. 9, n. 4, p. 275–312, 1982.

TARTAGLIONE, L. et al. Dereplication of Gambierdiscus balechii extract by LC-HRMS and in vitro assay: First description of a putative ciguatoxin and confirmation of 44-methylgambierone. **Chemosphere**, v. 319, 1 abr. 2023.

TAYLOR, A. J. et al. Microbes associated with spontaneous cacao fermentations - A systematic review and meta-analysis. **Current Research in Food Science**, v. 5, p. 1452–1464, jan. 2022.

THAKARE, R. et al. Antibiotics: past, present, and future. **Current opinion in microbiology**, v. 51, p. 72-80, 2019.

THIEMANN, T. Isolation of Phthalates and Terephthalates from Plant Material–Natural Products or Contaminants? **Open Chemistry Journal**, v. 8, n. 1, 2021.

TODESCHINI, R.; CONSONNI, V. **Handbook of molecular descriptors**. [s.l.] John Wiley & Sons, 2008.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of computational chemistry**, v. 31, n. 2, p. 455–461, 2010.

TSAVKELOVA, E. et al. Identification and functional characterization of indole-3-acetamide-mediated IAA biosynthesis in plant-associated Fusarium species. **Fungal Genetics and Biology**, v. 49, n. 1, p. 48–57, 2012.

TSLJCHIE, H. Effect of cacao husk extract on human immunodeficiency virus infection. **Letters in Applied Microbiology**, v. 13, n. 6, p. 251-254, 1991.

TSUGAWA, H. et al. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. **Nature Methods**, v. 12, n. 6, p. 523–526, 28 maio 2015.

UMEDA, M. et al. Preventive effects of tea and tea catechins against influenza and acute upper respiratory tract infections: A systematic review and meta-analysis. **European Journal of Nutrition**, v. 60, n. 8, p. 4189–4202, 2021a.

VALLI, M.; BOLZANI, V. S. Natural products: perspectives and challenges for use of Brazilian plant species in the bioeconomy. **Anais da Academia Brasileira de Ciências**, v. 91, 2019.

VAN DER SPOEL, D. et al. **GROMACS: Fast, flexible, and free**. **Journal of Computational Chemistry**, dez. 2005.

VERDONK, M. L. et al. Improved protein–ligand docking using GOLD. **Proteins: Structure, Function, and Bioinformatics**, v. 52, n. 4, p. 609–623, 2003.

VIEIRA, R. et al. Induction of metabolic variability of the endophytic fungus Xylaria sp. by OSMAC approach and experimental design. **Archives of Microbiology**, v. 203, n. 6, p. 3025–3032, 1 ago. 2021.

VIEIRA, R. et al. CHEIC: Chemical Image Classificator. An intelligent system for identification of volatiles compounds with potential for respiratory diseases using Deep Learning. **Expert Systems with Applications**, v. 234, p. 121178, dez. 2023.

VIEIRA, R.; ALVES DE SOUSA, K.; CASTRO-GAMBOA, I. **LUMIOS: Label Using Machine In Organic Samples-a software for dereplication, molecular docking, and combined machine and deep learning**. [s.l: s.n.].

VONRANKE, N. L. et al. Structure-activity relationship, molecular docking, and molecular dynamic studies of diterpenes from marine natural products with anti-HIV activity. **Journal of Biomolecular Structure and Dynamics**, v. 40, n. 7, p. 3185–3195, 2022.

WANG, G. et al. Trehalose and glucose levels regulate feeding behavior of the phloem-feeding insect, the pea aphid Acyrthosiphon pisum Harris. **Scientific Reports**, v. 11, n. 1, p. 15864, 2021.

WANG, W.; WANG, J.; KOLLMAN, P. A. What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? **Proteins: Structure, Function and Genetics**, v. 34, n. 3, p. 395–402, 15 fev. 1999.

WOSCHANK, M.; RAUCH, E.; ZSIFKOVITS, H. A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. **Sustainability (Switzerland)**, v. 12, n. 9, maio 2020.

WYNER, A. J. et al. Explaining the success of adaboost and random forests as interpolating classifiers. **The Journal of Machine Learning Research**, v. 18, n. 1, p. 1558–1590, 2017.

YAHYA, M.; GINTING, B.; SAIDI, N. In-Vitro Screenings for Biological and Antioxidant Activities of Water Extract from Theobroma cacao L. Pod Husk: Potential Utilization in Foods. **Molecules**, v. 26, n. 22, p. 6915, 2021.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, p. 611-629, 2018.

YAÑEZ, O. et al. Theobroma cacao L. compounds: Theoretical study and molecular modeling as inhibitors of main SARS-CoV-2 protease. **Biomedicine & Pharmacotherapy**, v. 140, p. 111764, 2021.

YANG, Z.; FANG, Y.; JI, H. Controlled release and enhanced antibacterial activity of salicylic acid by hydrogen bonding with chitosan. **Chinese Journal of Chemical Engineering**, v. 24, n. 3, p. 421–426, mar. 2016.

YARKONI, E.; BEKIERKUNST, A. Nonspecific resistance against infection with Salmonella typhi and Salmonella typhimurium induced in mice by cord factor (trehalose-6, 6'-dimycolate) and its analogues. **Infection and immunity**, v. 14, n. 5, p. 1125–1129, 1976.

YUAN, H. et al. The traditional medicine and modern medicine from natural products. **Molecules**, v. 21, n. 5, p. 559, 2016.

ZHANG, C. et al. Comparative Research on Network Intrusion Detection Methods Based on Machine Learning. **Computers & Security**, p. 102861, 2022.

ZHANG, H. et al. Organism-derived phthalate derivatives as bioactive natural products. **Journal of Environmental Science and Health, Part C**, v. 36, n. 3, p. 125–144, 2018.

ZHANG, J. et al. New lignans and their biological activities. **Chemistry & Biodiversity**, v. 11, n. 1, p. 1–54, 2014.

ZHANG, L. et al. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. **Drug discovery today**, v. 22, n. 11, p. 1680–1685, 2017.

ZHAO, J. et al. Plant-derived bioactive compounds produced by endophytic fungi. **Mini reviews in medicinal chemistry**, v. 11, n. 2, p. 159–168, 2011.

ZHENG, X.-Q. et al. Biosynthesis, accumulation and degradation of theobromine in developing Theobroma cacao fruits. **Journal of plant physiology**, v. 161, n. 4, p. 363-369, 2004.

ZHOU, D.-Y. et al. Proanthocyanidin from grape seed extract inhibits airway inflammation and remodeling in a murine model of chronic asthma. **Natural Product Communications**, v. 10, n. 2, p. 1934578X1501000210, 2015.

ZHU, W. et al. Anti-inflammatory and immunomodulatory effects of iridoid glycosides from Paederia scandens (LOUR.) MERRILL (Rubiaceae) on uric acid nephropathy rats. **Life Sciences**, v. 91, n. 11–12, p. 369–376, 5 out. 2012.

ZIĘBA, K.; MAKAREWICZ-WUJEC, M.; KOZŁOWSKA-WOJCIECHOWSKA, M. Cardioprotective Mechanisms of Cocoa. **Journal of the American College of Nutrition**, v. 38, n. 6, p. 564-575, 2019.

ZUMAETA, C. R. B. et al. Metabolomics during the spontaneous fermentation in cocoa (Theobroma cacao L.): An exploraty review. **Food Research International**, p. 112190, 2022.