



**UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

BEATRIZ LOBATO NOGUEIRA FARIAS

**Estudos *in silico* na busca de alcaloides inibidores do *Fusarium
oxysporum***

SÃO CRISTÓVÃO

2026

BEATRIZ LOBATO NOGUEIRA FARIAS

**ESTUDOS *IN SILICO* NA BUSCA DE ALCALOIDES
INIBIDORES DO *FUSARIUM OXYSPORUM***

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal de Sergipe como requisito à obtenção do grau de Mestre em Biotecnologia.

Orientador: Prof. Dr. Tiago Branquinho Oliveira

SÃO CRISTÓVÃO

2026

BEATRIZ LOBATO NOGUEIRA FARIAS

**ESTUDOS *IN SILICO* NA BUSCA DE ALCALOIDES
INIBIDORES DO *FUSARIUM OXYSPORUM***

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal de Sergipe como requisito à obtenção do grau de Mestre em Biotecnologia.

Orientador: Prof. Dr. Tiago Branquinho Oliveira

1º Examinador: Prof. Dr. Marcus Vinicius de Aragão Batista

2º Examinador: Prof. Dr. José Diogo de Lisboa Dutra

Agradecimentos

Agradeço aos meus pais, Denilson e Oscarina, a minha família, em especial minha querida avó Valquíria e Tio João por sempre acreditarem em mim a seguir meus sonhos, aos meus amigos, e todos que me apoiaram durante esses anos morando em Aracaju, estão todos em meu coração e serei eternamente grata;

Agradeço ao meu orientador Prof. Dr. Tiago Branquinho Oliveira, por todos os momentos de paciência, prontidão, disponibilidade em sempre ajudar e esclarecer quaisquer dúvidas, onde todos os encontros eram realizados com muito conhecimento e aprendizado que levarei para sempre;

Agradeço ao grupo de pesquisa do Laboratório de Alimentos e Bebidas (LAAB) e ao Departamento de Farmácia da UFS, por toda colaboração e auxílio durante o desenvolvimento deste trabalho, aos colegas do grupo de pesquisa *Computer Aided Science* (CASE) pela parceria e troca de conhecimento;

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro;

Agradeço à Universidade Federal de Sergipe (UFS), pelo espaço disponibilizado, e em menção especial, ao Programa de Pós-Graduação em Biotecnologia (PROBIO), pela oportunidade de desenvolvimento acadêmico.

Beatriz Lobato Nogueira Farias, Prof. Dr. Tiago Branquinho Oliveira, Estudos *in silico* na busca de alcaloides inibidores do *Fusarium oxysporum*, Dissertação de Mestrado, Programa de Pós-Graduação em Biotecnologia (PROBIO), 2026.

RESUMO:

O fungo *Fusarium oxysporum* é um fitopatógeno amplamente distribuído, responsável por prejuízos agrícolas significativos e capaz de atuar como agente oportunista em infecções humanas. Métodos de controle, são comumente empregados a aplicação de fungicidas e rotação de culturas, entretanto, o micro-organismo tem apresentado mecanismos de resistência a diferentes antifúngicos, limitando a eficácia dos tratamentos disponíveis. Nesse contexto, torna-se essencial a busca por novas substâncias com propriedades inibitórias, visando o desenvolvimento de alternativas terapêuticas e o aprimoramento das estratégias de combate ao patógeno. O presente estudo tem como objetivo identificar alcaloides de potencial inibitório contra *F. oxysporum* por meio de abordagem *in silico*. Um banco de dados proveniente do ChEMBL foi selecionado, contendo estruturas previamente avaliadas em ensaios de concentração inibitória mínima (MIC). Em seguida, estruturas semelhantes a alcaloides foram analisadas, com auxílio das ferramentas algoritmo genético e *random forest* 6 descritores moleculares foram selecionados para classe AB. O modelo classificatório desenvolvido apresentou parâmetros estatísticos consistentes em valores de kappa de cohen (interna = 0,935; cruzada = 0,751; externa = 0,740), indicando previsibilidade e robustez. Em seguida, os modelos foram avaliados por meio de validação estatística, incluindo a matriz de confusão, estatísticas de classe (recall, precisão, sensibilidade e especificidade), estatísticas gerais (previsão e erro geral, kappa de cohen, estruturas classificadas corretamente e incorretamente) e curva ROC. Posteriormente, realizada a triagem virtual, foram selecionadas estruturas de 70 alcaloides ativos com potencial antifúngico a partir de 85% de confiabilidade. Os resultados evidenciam a aplicabilidade da quimioinformática no suporte à validação experimental e no estudo de novas estruturas bioativas.

Palavras chave: Alcaloides, Triagem virtual, Fusarium oxysporum, QSAR.

Beatriz Lobato Nogueira Farias, Prof. Dr. Tiago Branquinho Oliveira, *In silico* studies in the search for alkaloids that inhibit *Fusarium oxysporum*, Graduate Program in Biotechnology (PROBIO), 2026.

ABSTRACT:

The fungus *Fusarium oxysporum* is a widely distributed phytopathogen responsible for significant agricultural losses and capable of acting as an opportunistic agent in human infections. Control methods commonly employed include the application of fungicides and crop rotation; however, the microorganism has shown resistance mechanisms to different antifungals, limiting the effectiveness of available treatments. In this context, the search for new substances with inhibitory properties becomes essential, aiming at the development of therapeutic alternatives and the improvement of strategies to combat the pathogen. This study aims to identify alkaloids with inhibitory potential against *F. oxysporum* through an *in silico* approach. A database from ChEMBL was selected, containing structures previously evaluated in minimum inhibitory concentration (MIC) assays. Subsequently, alkaloid-like structures were analyzed using genetic algorithm and random forest tools; 6 molecular descriptors were selected for class AB. The developed classification model presented consistent statistical parameters in Cohen's kappa values (internal = 0.935; cross = 0.751; external = 0.740), indicating predictability and robustness. Subsequently, the models were evaluated through statistical validation, including the confusion matrix, class statistics (recall, precision, sensitivity, and specificity), general statistics (prediction and overall error, Cohen's kappa, correctly and incorrectly classified structures), and ROC curve. Following virtual screening, structures of 70 active alkaloids with antifungal potential were selected based on 85% reliability. The results highlight the applicability of chemoinformatics in supporting experimental validation and the study of new bioactive structures.

Keywords: Alkaloids, Virtual Screening, Fusarium oxysporum, QSAR.

SUMÁRIO

Resumo	5
Abstract	6
Lista de ilustrações	8
Lista de tabelas	9
Abreviaturas e siglas	10
1. Introdução	11
1.1 <i>Fusarium oxysporum</i>	11
1.1.1 Controle e Resistência Microbiana	14
1.2 Alcaloides como promissores inibidores fúngicos	14
1.3 Quimioinformática no planejamento de produtos fitofarmacêuticos.....	16
1.3.1 QSAR.....	18
1.4 Prospecção <i>in silico</i> de alcaloides antifúngicos contra <i>Fusarium oxysporum</i>	20
2. Justificativa	23
3. Objetivos	24
3.1 Objetivo Geral.....	24
3.2 Objetivos Específicos.....	24
4. Metodologia	25
4.1 <i>Softwares</i>	25
4.2 Métodos.....	26
4.2.1 Modelagem QSAR para <i>F. oxysporum</i>	26
4.2.2 Validação dos modelos	32
4.2.3 Triagem virtual.....	37
5. Resultados e discussão	39
5.1 Modelagem QSAR para <i>F. oxysporum</i>	39
5.1.1 Construção dos modelos	39
5.2 Validação dos modelos	42
5.2.1 Validação Interna	43
5.2.2 Validação Cruzada	46
5.2.3 Validação Externa	50
5.3 Triagem Virtual.....	53
5.3.1 Análise da triagem virtual.....	56
6. Conclusão	59
Referências bibliográficas	60

LISTA DE ILUSTRAÇÕES

Lista de Quadros

- Quadro 1:** Exemplos de para formas especiais de *Fusarium oxysporum* em diversas plantas.. 12
- Quadro 2:** Descritores moleculares e suas quantidades calculadas pelo Alvadesk para pMIC. 27
- Quadro 3:** Matriz de confusão para uma classificação binária..... 34

Lista de Figuras

- Figura 1:** Estrutura dimensional (2D) da Quinina. 15
- Figura 2:** Estrutura dimensional (2D) da Mescalina..... 15
- Figura 3:** Estrutura dimensional (2D) da Coniina. 15
- Figura 4:** Gráfico ROC com cinco classificadores. 36
- Figura 5:** Curva ROC da validação interna para classe A. 45
- Figura 6:** Curva ROC da validação cruzada para classe A..... 49
- Figura 7:** Curva ROC da validação externa para classe A. 53

LISTA DE TABELAS

Tabela 1: Valores para interpretação do coeficiente Kappa de Cohen.	33
Tabela 2: Interpretações dos valores de AUC.	37
Tabela 3: Importância relativa dos descritores selecionados para classe AB.....	39
Tabela 4: Parâmetros avaliados na validação interna do modelo QSAR desenvolvido.	43
Tabela 5: Parâmetros avaliados na validação cruzada do modelo QSAR desenvolvido.....	47
Tabela 6: Parâmetros avaliados para validação externa do modelo QSAR desenvolvido.	51
Tabela 7: 10 estruturas selecionadas a partir de maior confiabilidade de atividade na Triagem Virtual.	54

ABREVIATURAS E SIGLAS

AG – Algoritmo Genético

CADD – Desenho de Fármacos Assistido por Computador (*Computer Aided Drug Design*)

LBDD – Planejamento de fármacos baseado na estrutura do ligante (*Ligand-Based Drug Design*)

MIC – Concentração Inibitória Mínima

ODS – Objetivos de Desenvolvimento Sustentável

PPPs – Produtos de proteção de plantas (*Plant Protection Products*)

QSAR – Relação Estrutura-Atividade Quantitativa (*Quantitative Structure-Activity Relationship*)

QSPR – Relação Estrutura-Propriedade Quantitativa (*Quantitative Structure-Property Relationship*)

RF – Floresta Randômica (*Random Forest*)

SBDD – Planejamento de fármacos baseado na estrutura do receptor (*Structure-Based Drug Design*)

TFP – Taxa de Falsos Positivos

TV – Triagem Virtual (*Virtual Screening*)

TVP – Taxa de Verdadeiros Positivos

1. INTRODUÇÃO

1.1 *Fusarium oxysporum*

Espécies do gênero *Fusarium* estão entre os fungos mais amplamente distribuídos no meio ambiente, sendo comumente encontradas no solo e em matéria orgânica em decomposição (Román-Montes *et al.*, 2025). Esses fungos podem atuar como endófitos, estabelecendo relações simbióticas com plantas durante parte de seu ciclo de vida sem causar danos aparentes, ou como saprófitas, obtendo nutrientes a partir da degradação de matéria orgânica (Coleman, 2016). O gênero apresenta espécies patogênicas e não patogênicas de grande relevância agrícola, veterinária e médica, caracterizando-se pela elevada capacidade metabólica e pela produção de metabólitos secundários bioativos, incluindo compostos antifúngicos, antibacterianos e citotóxicos (Coleman, 2016; Gordon, 2025; Manici; Caputo; Saccà, 2017). As espécies de *Fusarium* produzem diferentes tipos de esporos — macroconídios, microconídios e clamidósporos — formados a partir de hifas especializadas (Gordon, 2025) e micotoxinas — deoxinivalenol e fumonisinas — perigosas para a segurança alimentar, representando ameaças à saúde humana (Munkvold, 2017).

A fusariose, também conhecida como murcha-de-*Fusarium* (*Fusarium wilt*), é uma das principais doenças de plantas de interesse agrônomico, afetando culturas economicamente relevantes como banana, algodão, soja, melão e cebola (Edel-Hermann; Lecomte, 2019). Causada principalmente por *Fusarium oxysporum*, a doença manifesta-se por sintomas característicos como murcha progressiva, amarelecimento foliar generalizado, escurecimento vascular e, em alguns casos, murcha unilateral das folhas (Lopes; Reis, 2022). Internamente, podem ser observadas descolorações vasculares no pseudocaule, além de podridão de sementes e frutos, especialmente em áreas de bananicultura (Fernandes *et al.*, 2008; Michereff *et al.*, 2005).

A ampla gama de hospedeiros de *F. oxysporum* reflete a elevada diversidade de suas formas especiais (*formae speciales*), conceito introduzido por Eriksson em 1894 para descrever a especialização patogênica em fungos morfológicamente semelhantes. Posteriormente, Snyder e Hansen (1940) relataram aproximadamente 25 formas biológicas associadas a diferentes hospedeiros, alguns exemplos podem ser observados no Quadro 1. Avanços taxonômicos e fitopatológicos ocorridos entre 1965 e 2012 permitiram a reclassificação dessas formas biológicas em formas especiais, algumas das quais subdivididas em raças, em função do conhecimento sobre genes de resistência das plantas hospedeiras. Atualmente, são reconhecidas e bem documentadas 106 formas especiais de *F. oxysporum*, associadas a 45 famílias botânicas,

enquanto outras permanecem com caracterização limitada (Edel-Hermann; Lecomte, 2019). A diversidade desse patógeno é evidenciada pela ampla gama de plantas hospedeiras, incluindo magnoliopsidas (anteriormente mencionadas como dicotiledôneas), como feijão e tomate, e monocotiledôneas, como banana e palmeiras, além de espécies com ciclos de vida anuais ou perenes, terrestres ou aquáticas (Edel-Hermann; Lecomte, 2019).

Quadro 1: Exemplos de para formas especiais de *Fusarium oxysporum* em diversas plantas.

Formas especiais	Exemplo de planta hospedeira
<i>cepae</i>	Cebola
<i>conglutinans</i>	Repolho
<i>cubense</i>	Banana
<i>cucumerinum</i>	Pepino
<i>lycopersici</i>	Tomate
<i>melonis</i>	Melão
<i>Niveum</i>	Melancia
<i>psidii</i>	Goiaba

Fonte: Adaptado de Lievens *et al.*, 2009.

A importância fitossanitária de *Fusarium* no Brasil é evidenciada pela ampla distribuição de suas formas especiais em diferentes regiões produtoras, associada à economia agrícola nacional. Entre as patologias causadas por espécies de *Fusarium*, a fusariose se destaca por seu impacto econômico em diversas culturas de grande relevância, como banana (*Musa* spp.) e tomate (*Solanum lycopersicum*), podendo reduzir significativamente produtividade e qualidade da produção em escala comercial.

Pesquisadores da Embrapa Amazônia Ocidental (AM), Embrapa Roraima e *Bioversity Internacional* em parceria com a Superintendência Federal de Agricultura no Amazonas (SF/AM) produziram um relatório técnico na tentativa de antecipar o problema proveniente da raça 4 tropical (TR4) do *Fusarium oxysporum* f. sp. *Cubense*, ausente no Brasil em 2020 e para a qual não haviam sido relatados cultivares resistentes aos bananais. Como também, relatam que a TR4

estaria disseminada na Austrália, China, Filipinas, Indonésia, Laos, Líbano, Malásia, Moçambique, Myanmar, Omã, Paquistão, Taiwan e Vietnã e por final constatada na Colômbia, colocando em risco a bananicultura nacional.

Registros do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) em 2021 indicaram a ocorrência da fusariose em diferentes estados brasileiros, com variações na frequência e severidade dependendo do agente causal, das condições edafoclimáticas e da susceptibilidade das cultivares. Na cultura bananeira, a fusariose causada por *Fusarium oxysporum* f. sp. *cubense* é reconhecida como uma das principais limitações fitossanitárias no país. Estudos realizados no estado de São Paulo, nas regiões do Vale do Ribeira e de São Bento do Sapucaí, documentaram ocorrências significativas da doença, evidenciando que diferentes cultivares apresentam níveis de susceptibilidade variáveis ao agente patogênico. Nesses estudos, a fusariose foi caracterizada como uma das principais doenças que reduzem o vigor das plantas e a produtividade das lavouras bananais, especialmente em áreas com histórico de cultivo contínuo (Teixeira *et al.*, 2021).

Segundo dados oficiais da Produção Agrícola Municipal (PAM), em 2023 o Brasil se destacou como um dos maiores produtores mundiais de banana, com geração de valor estimada em R\$ 13,8 bilhões, posicionando-se entre os principais produtores das regiões Nordeste e Sudeste. A cultura do tomate também teve destaque, com rendimento aproximado de R\$ 10 bilhões em cerca de 59 mil hectares cultivados no mesmo ano (IBGE, 2023). A elevada participação socioeconômica dessas culturas também se reflete na importância de manter a sanidade das plantações diante de doenças fúngicas de alta prevalência, como a fusariose.

Em cultivos de tomate, pesquisas recentes também relatam a presença de fusariose em diferentes regiões do país. Por exemplo, em áreas produtoras do estado do Pará, nos municípios de Santarém e Belterra, foram isolados agentes patogênicos pertencentes a *F. solani* e *F. oxysporum*, associados à manifestação de sintomas de murcha e declínio vegetativo das plantas, destacando a ocorrência de múltiplas espécies de *Fusarium* como agentes fitopatogênicos em horticulturas brasileiras (Severo *et al.*, 2024). Esses achados reforçam a necessidade de monitoramento fitossanitário contínuo, identificação de espécies e formas especiais envolvidas, bem como a adoção de estratégias integradas de manejo que considerem resistência genética, práticas culturais adequadas e controle biológico. Essa distribuição geográfica revela tanto a adaptação ecológica de *Fusarium* quanto a necessidade de estudos que considerem a interação entre o patógeno, o hospedeiro e o ambiente no contexto brasileiro.

1.1.1 Controle e Resistência Microbiana

Ao longo das últimas décadas, o controle químico a partir da utilização de fungicidas reflete em um dos principais procedimentos no controle de doenças de plantas. Apesar da comodidade no momento da aplicação e seus resultados imediatos, seu uso contínuo apresenta riscos em relação ao controle e a eficiência do método utilizado previamente (Ghini; Kimati, 2000). Revelando a necessidade de fungicidas direcionados a locais de ação diferentes devido a resistência criada por aplicações contínuas, ameaçando diversas culturas com patógenos vegetais (Yin *et al.*, 2023). Casos de resistência de *Fusarium* spp. a fungicidas foram relatadas em diversas regiões do mundo, onde sua persistência varia conforme espécie, técnicas agrícolas aplicadas, padrões na utilização de fungicidas e localização geográfica (Bai; Su; Cai, 2018; De Chaves *et al.*, 2022; Naqvi *et al.*, 2025).

Espécies de *Fusarium* demonstram resistência a diversas classes de fungicidas, como: azólicos, benzimidazóis, estrobilurinas e triazóis (Yin *et al.*, 2023; Zhao; He; Wang, 2021). Segundo Johns *et al.* (2022), a identificação de padrões de resistência colabora ao desenvolvimento de estratégias específicas, na redução do impacto de doenças associadas ao *Fusarium* e à contaminação por micotoxinas.

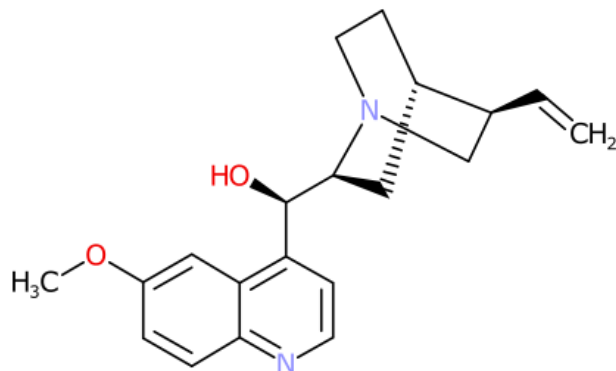
1.2 Alcaloides como promissores inibidores fúngicos

Os alcaloides constituem um importante grupo de metabólitos secundários, sendo estruturas orgânicas amplamente distribuídas em fungos, bactérias, plantas, animais e organismos marinhos. Caracterizam-se por elevada diversidade estrutural e pela presença de pelo menos um átomo de nitrogênio em sua estrutura química, geralmente em estado básico, não pertencente a ligações do tipo amida ou peptídica. Esses compostos podem ser encontrados em diferentes compartimentos celulares, como mitocôndrias, vesículas, cloroplastos e vacúolos (Gutiérrez-Grijalva *et al.*, 2020).

Os estudos científicos sobre alcaloides tiveram início em 1806, com o isolamento da morfina a partir da papoula (*Papaver somniferum*) pelo farmacêutico alemão Friedrich Wilhelm Adam Sertürner, considerado um marco histórico no desenvolvimento da química de produtos naturais (Goerig; Schulte am Esch, 1991). Com base em sua estrutura molecular e nas vias biossintéticas envolvidas, os alcaloides são tradicionalmente classificados em três subgrupos:

(a) alcaloides verdadeiros ou heterocíclicos, como a quinina (Figura 1);

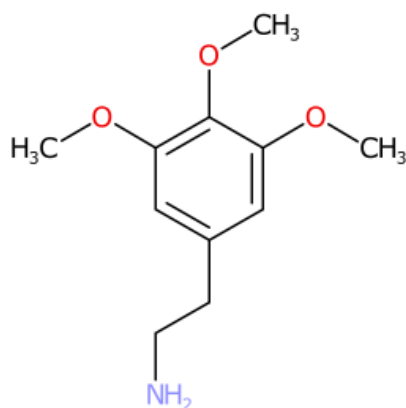
Figura 1: Estrutura dimensional (2D) da Quinina.



Fonte: Autor.

(b) protoalcaloides, de estrutura não heterocíclica, como a mescalina (Figura 2);

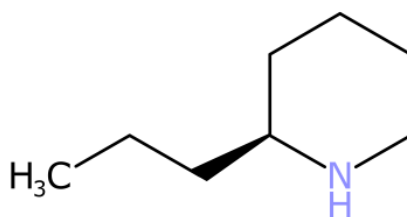
Figura 2: Estrutura dimensional (2D) da Mescalina.



Fonte: Autor.

e (c) pseudoalcaloides, a exemplo da coniina (Figura 3).

Figura 3: Estrutura dimensional (2D) da Coniina.



Fonte: Autor.

Além de desempenharem papel fundamental na defesa química de plantas, vertebrados e microrganismos contra agentes externos, como predadores e patógenos (Goyal, 2013), os alcaloides apresentam ampla gama de atividades biológicas. Entre essas, destacam-se as atividades antifúngica e antibacteriana, exemplificadas pela berberina; antiparasitária, como a quinina; e citotóxica, atribuída a compostos como vimblastina e vincristina, amplamente utilizadas na terapêutica antineoplásica (Simões *et al.*, 2016).

Estudos recentes têm evidenciado o potencial dos alcaloides no controle de fungos fitopatogênicos, despertando crescente interesse da comunidade científica (Bhambhani; Kondhare; Giri, 2021; Wu *et al.*, 2021). Alcaloides β -carbolínicos isolados de *Peganum harmala* demonstraram potente inibição do crescimento micelial de *Fusarium oxysporum*, promovendo alterações morfológicas e induzindo o acúmulo de espécies reativas de oxigênio (ROS) em células fúngicas (Zhu *et al.*, 2022). De forma semelhante, extratos alcaloidais de *Lupinus mirabilis* apresentaram atividade fungistática significativa, sendo identificados alcaloides do tipo quinolizidínico com potencial aplicação no desenvolvimento de fitofármacos (Bernal; Coy-Barrera, 2022). Esses resultados reforçam a relevância da investigação fitoquímica e da avaliação biológica de alcaloides como estratégia promissora para o controle de fitopatógenos de importância agrícola.

Diante da relevância agronômica de *Fusarium oxysporum* e das limitações associadas ao controle químico convencional, torna-se evidente a necessidade de estratégias alternativas que aliem eficácia antifúngica, seletividade e menor impacto ambiental. Nesse contexto, compostos naturais, especialmente alcaloides, despontam como fontes promissoras de novas moléculas bioativas.

1.3 Quimioinformática no planejamento de produtos fitofarmacêuticos

O termo quimioinformática foi introduzido em 1998 por Frank Brown, sendo definido como “uma combinação de informações destinada a transformar dados em informação e informação em conhecimento, com o objetivo de subsidiar decisões mais rápidas e eficazes nos processos de identificação e otimização de estruturas líderes”. Trata-se de uma ciência interdisciplinar que integra métodos das ciências da computação e da informação para a resolução de problemas relacionados à química, especialmente no tratamento e interpretação de grandes volumes de dados químicos (Gasteiger, 2003). Atualmente, a quimioinformática compreende três eixos principais: (i) consolidação, visualização, manipulação e processamento de estruturas

químicas; (ii) organização e curadoria de bases de dados contendo informações estruturais e físico-químicas; e (iii) estudos das relações quantitativas entre estrutura e atividade (QSAR) ou entre estrutura e propriedade (QSPR) (Willett, 2011).

No contexto do desenvolvimento de produtos fitofarmacêuticos, a quimioinformática desempenha papel fundamental no planejamento racional de substâncias bioativas, contribuindo para a predição de propriedades toxicológicas, ambientais e de eficácia biológica por meio de abordagens *in silico*. Em especial, no cenário europeu, essas ferramentas têm sido amplamente empregadas como suporte aos processos de avaliação de risco e tomada de decisão regulatória, reduzindo a dependência de ensaios experimentais extensivos (European Food Safety Authority, 2013; More *et al.*, 2021; Carrasco Cabrera *et al.*, 2023).

Os pesticidas, também denominados produtos de proteção de plantas (*plant protection products* – PPPs), são amplamente utilizados na agricultura, silvicultura e jardinagem urbana com o objetivo de proteger as culturas contra pragas, como insetos e plantas daninhas, e contra agentes fitopatogênicos, a exemplo de fungos. A aplicação desses produtos visa preservar a sanidade vegetal, minimizar perdas produtivas e garantir a estabilidade dos sistemas agrícolas, sendo, portanto, um componente estratégico para a segurança alimentar. No entanto, seu uso também demanda rigorosa avaliação de riscos ambientais e toxicológicos, aspecto no qual ferramentas quimioinformáticas assumem relevância crescente (Frische *et al.*, 2018).

Os estudos de Rodrigues *et al.* (2022) e Mukherjee, Kumar e Roy (2022) exemplificam aplicações complementares da modelagem QSAR no contexto de produtos de proteção de plantas, tanto no desenvolvimento de novas moléculas quanto na avaliação de segurança ambiental. Rodrigues *et al.* (2022) empregaram modelos QSAR guiados por aprendizado de máquina para propor derivados de tioureias com potencial herbicida mais eficaz e seguro, utilizando dados experimentais de inibição do crescimento radicular de *Brassica napus*, o que confere validação biológica direta às correlações estrutura–atividade estabelecidas. Por outro lado, Mukherjee, Kumar e Roy (2022) desenvolveram modelos 2D-QSAR voltados à predição da toxicidade aguda por contato de pesticidas em abelhas (*Apis mellifera*), fundamentando-se em dados experimentais previamente reportados na literatura e em bases regulatórias, com validação estatística rigorosa segundo recomendações da OCDE, porém sem a realização de novos ensaios *in vivo*. Em conjunto, esses trabalhos demonstram o potencial da abordagem QSAR tanto como ferramenta de apoio ao desenho racional de novos agroquímicos quanto como estratégia robusta

para avaliação de riscos toxicológicos e ambientais, contribuindo para o desenvolvimento de produtos mais seguros e sustentáveis.

De forma aplicada, a área abriga estratégias no planejamento de bioativos auxiliado por computador (CADD) que utilizam combinações de diversas ferramentas computacionais com propósito de identificar, desenvolver ou otimizar estruturas promissoras a futuros fitofarmacêuticos, auxiliando no longo e custoso processo de descoberta de novas substâncias. As ferramentas utilizadas são baseadas em estruturas (SBDD) ou em ligantes (LBDD). Para o SBDD, temos a docagem molecular utilizada na predição de orientações entre complexos ligante-receptor. No LBDD, temos os estudos de QSAR (Rodrigues *et al.*, 2012).

1.3.1 QSAR

Para identificar novas moléculas com potencial terapêutico, o design computacional de fármacos, especialmente por meio de análises de *quantitative structure–activity relationship* (QSAR), constitui uma abordagem viável e historicamente consolidada. A metodologia QSAR foi proposta por Hansch e Fujita em 1964 e permanece como uma técnica eficiente para o desenvolvimento de modelos matemáticos capazes de correlacionar descritores moleculares com atividades biológicas relevantes (Neves *et al.*, 2018). Segundo Iman *et al.* (2014), essa ferramenta permite compreender a relação entre parâmetros físico-químicos de uma estrutura química e sua atividade biológica, o que auxilia pesquisadores a reduzir o tempo de desenvolvimento e minimizar custos nos estágios iniciais de pesquisa e síntese de compostos.

O principal desafio na modelagem QSAR é a construção de modelos precisos e confiáveis, dado o elevado grau de complexidade na relação entre estrutura molecular e atividade biológica. Nos últimos anos, observou-se a integração de técnicas de aprendizado de máquina — como regressão, classificação e redes neurais — ampliando a capacidade preditiva desses modelos (Bao *et al.*, 2023; Dai *et al.*, 2022). Além disso, a aplicação de QSAR tem se expandido para a predição de propriedades toxicológicas e ambientais de compostos, incluindo pesticidas e contaminantes emergentes, reforçando a importância dessa abordagem para estudos de segurança química (Mukherjee; Kumar; Roy, 2022).

A técnica QSAR baseia-se em dois princípios fundamentais: compostos com estruturas semelhantes tendem a apresentar comportamentos biológicos comparáveis sob condições ambientais semelhantes, e diferenças nas atividades podem ser relacionadas a variações na composição ou na estrutura molecular. As variáveis preditoras são denominadas *descritores* (por

exemplo, propriedades físico-químicas ou estruturais), enquanto as variáveis resposta (como reatividade, bioatividade ou toxicidade) são referidas como *atividades* (Chen *et al.*, 2015).

Diversos algoritmos de aprendizado de máquina são empregados na modelagem QSAR, incluindo árvores de decisão (DT), regressão linear múltipla (MLR), máquina de vetores de suporte (SVM) e redes neurais perceptron multicamadas (MLP) (Damale *et al.*, 2014).

Duchowicz *et al.* (2007) desenvolveram modelos QSAR utilizando centenas de descritores topológicos, geométricos e eletrônicos na correlação de estruturas e atividades antifúngicas de derivados heterocíclicos. Como também, Andrade *et al.* (2012) aplicaram modelos QSAR 2D com descritores topológicos e fragmentos moleculares para modelos antifúngicos de derivados azólicos contra *Candida albicans*, na importância ao entendimento da atividade biológica relacionada a descritores topológicos. Tetz *et al.* (2023) em estudos com compostos aminados como M451 (derivado de 1,6-diamino-hexano) mostraram atividade antifúngica significativa contra diferentes fungos fitopatogênicos, incluindo *Fusarium oxysporum*, com valores de concentração fungicida mínima de 4 a 512 µg/mL com variações de tempo de exposição de 5 minutos a 24 horas. Ensaio de germinação de conídios induziram 99% de inibição ao *F. oxysporum* em 5 minutos de exposição.

Estudos *in silico* realizados na Universidade Federal de Sergipe (UFS) demonstram aplicações modernas de QSAR em diferentes alvos terapêuticos. Fernandes *et al.* (2025) realizaram uma análise de flavonoides com potencial antimalárico por meio de QSAR, acoplamento molecular e simulações de dinâmica molecular na predição de atividade contra enzimas de *Plasmodium falciparum*. Ferreira *et al.*, (2021) combinaram uma biblioteca de produtos naturais e triagem virtual baseada em QSAR, priorizando compostos naturais e derivados com atividade potencial antimalárica, validado em teste experimental contra *Plasmodium falciparum* sensíveis a cloroquina, confirmando as seleções *in silico* de duas lactonas sesquiterpênicas (LDT-597 e LDT-598). Os estudos empregam a utilização da inteligência artificial para prever atividades de estruturas e elucidar suas relações estrutura-atividade, visando o desenvolvimento de novas alternativas terapêuticas.

De forma análoga, Paixão *et al.* (2025) empregaram abordagens integradas de QSAR e docagem molecular para a identificação de inibidores de SmTGR, uma enzima essencial para *Schistosoma mansoni*. Em trabalhos publicados na literatura científica sobre o desenvolvimento de modelos QSAR para inibir SmTGR, modelos preditivos foram utilizados para realizar triagens virtuais seguidas de ensaios *in vitro* que avaliaram a eficácia dos compostos contra larvas e

vermes adultos do parasita, confirmando a atividade biológica prevista pelo modelo QSAR e destacando compostos com potencial esquistossomicida (em baixos micromolares) (Neves *et al.*, 2016). Esses exemplos ilustram como estudos QSAR podem ser integrados a validações experimentais, fortalecendo a utilidade prática da abordagem *in silico* no processo de descoberta de novas drogas.

No que concerne à validação estatística de modelos QSAR, diferentes estratégias são empregadas para avaliar sua robustez preditiva, incluindo validação interna (*training set*), validação cruzada e validação externa (*test set*). O particionamento adequado dos conjuntos de treinamento e teste, junto a técnicas complementares como *bootstrapping*, é essencial para assegurar que os modelos QSAR sejam confiáveis e gerem previsões úteis e generalizáveis (Panov; Džeroski, 2007; Parvandehe *et al.*, 2020).

De maneira geral, modelos QSAR são amplamente utilizados na predição de propriedades biológicas de novas substâncias, sendo valorizados por sua capacidade de gerar estimativas com alta taxa de acerto e produtividade, o que os torna ferramentas valiosas no contexto da descoberta de fármacos e fitofarmacêuticos (Carpenter *et al.*, 2018; Neves *et al.*, 2018). O avanço das abordagens *in silico*, em particular da quimioinformática e da modelagem QSAR, tem permitido acelerar a identificação de candidatos bioativos, reduzindo custos experimentais e ampliando a racionalidade no processo de descoberta de novos agentes antifúngicos. Contudo, a aplicação sistemática dessas ferramentas voltada à prospecção de alcaloides com atividade específica contra *F. oxysporum* ainda permanece pouco explorada, configurando um campo relevante para investigação científica.

1.4 Prospecção *in silico* de alcaloides antifúngicos contra *Fusarium oxysporum*

A aplicação de métodos computacionais na busca por novos antifúngicos tem se consolidado como uma estratégia racional e de alta produtividade, especialmente quando integrada a bancos de dados de produtos naturais (Wang *et al.*, 2025); Neves *et al.*, 2018). No âmbito dos alcaloides, diversas classes como quinolizidínicos, β -carbolínicos e isoquinolínicos, têm sido investigadas quanto ao seu potencial inibitório contra fungos fitopatogênicos, incluindo espécies de *Fusarium* (Bernal; Coy-Barrera, 2022; Li; Zhao; Zhang, 2021; Zhu; Zhao; Wang, 2022). Contudo, a incorporação de ferramentas quimioinformáticas, em especial a modelagem por relações quantitativas estrutura-atividade (QSAR), ainda está em fase inicial quando o alvo específico é *F. oxysporum*.

Uma das abordagens diretamente relacionadas a esta dissertação foi apresentada por Bernal e Coy-Barrera (2019), durante o Brazilian Conference on Natural Products. Os autores construíram um modelo de QSAR-3D a partir de 76 compostos com atividade antifúngica conhecida contra *F. oxysporum* (extraídos do ChEMBL) e 21 alcaloides quinolizidínicos isolados de leguminosas dos gêneros *Lupinus*, *Genista* e *Ulex*. O modelo foi gerado por regressão linear com descritores Volsurf+, identificou que a presença do anel 2-piridona e derivados de citisina são determinantes para a atividade, evidenciando que alcaloides dessa classe possuem potencial preditivo e podem ser racionalmente explorados por métodos *in silico*. Os autores aplicaram o modelo validado externamente para prever a atividade de 100 alcaloides quinolizidínicos ainda não testados, apontando candidatos com IC_{50} predito entre 1 e 50 $\mu\text{g/mL}$.

Recentemente, o mesmo grupo de pesquisa validou experimentalmente o potencial antifúngico de alcaloides quinolizidínicos contra *F. oxysporum* (Cely-Veloza *et al.*, 2023). Nesse estudo, vinte alcaloides pertencentes a seis classes estruturais: lupanina, esparteína, lupinina, citisina, tetra-hidrocitisina e matrina. Foram isolados de *Lupinus polyphyllus*, *L. mutabilis* e *Genista monspessulana* cultivados em casa de vegetação e tiveram sua atividade antifúngica determinada por meio de ensaios *in vitro* (Cely-Veloza *et al.*, 2023). Os compostos esparteína, aflina, citisina e matrina apresentaram os melhores perfis inibitórios, com valores de IC_{50} entre 7,2 e 16,5 μM , sendo classificados como fungicidas e fungistáticos (Cely-Veloza *et al.*, 2023). Adicionalmente, os autores estabeleceram relações de estrutura-atividade (SAR) que indicam requisitos estruturais específicos, como a presença de sistemas bicíclicos fundidos e grupos carbonila em posições estratégicas, para a otimização da potência antifúngica (Cely-Veloza *et al.*, 2023).

Em conjunto, os trabalhos de Bernal e Coy-Barrera (2019, 2023) constituem, até o momento, a iniciativa mais sistemática na prospecção de alcaloides quinolizidínicos contra *F. oxysporum*, integrando modelagem preditiva e validação experimental. Contudo, tais estudos mencionam a apenas uma única classe de alcaloides (quinolizidínicos), não exploram a diversidade estrutural de outros alcaloides (isoquinolínicos, indólicos, esteroidais) e não empregam validação externa com compostos alheios aos conjuntos de treino. Sendo estas, lacunas que a presente pesquisa pretende endereçar por meio da modelagem QSAR aplicada a um espectro mais amplo de alcaloides.

Em paralelo, Li, Zhao e Zhang, (2021) isolaram o alcaloide canthin-6-ona de *Ailanthus altissima* e demonstraram por meio de ensaios proteômicos, que este composto inibe *F.*

oxysporum f. sp. *cucumerinum* ao interferir nas vias de biossíntese de aminoácidos e no metabolismo de nitrogênio. Embora o foco do trabalho não fosse a modelagem QSAR, ele fornece alvos metabólicos validados experimentalmente que podem subsidiar futuras análises estrutura-atividade. De forma análoga, Zhu *et al.* (2022) evidenciaram que alcaloides β -carbolínicos de *Peganum harmala* promovem acúmulo de espécies reativas de oxigênio (ROS) em *F. oxysporum*, associando a atividade antifúngica a características estruturais específicas, uma informação que poderia ser explorada em modelos QSAR.

No campo da modelagem preditiva aplicada a pesticidas e antifúngicos, Rodrigues *et al.* (2022) e Mukherjee, Kumar & Roy (2022) exemplificam o uso de QSAR combinado com aprendizado de máquina para o desenvolvimento de herbicidas mais seguros e para predição de toxicidade ambiental. Esses estudos demonstram a maturidade das ferramentas QSAR para o planejamento de agroquímicos, mas nenhum deles aborda especificamente alcaloides ou o patógeno *F. oxysporum*, o que mostra uma lacuna clara na literatura.

Assim como, Wang *et al.* (2025) empregaram um pipeline de machine learning (Random Forest, SVM e redes neurais) para rastrear inibidores contra *F. solani* e validaram experimentalmente que a cefarantina (um alcaloide bisbenzilisquinolínico) também é ativa contra *F. oxysporum*. O estudo utiliza descritores moleculares (RDKit) e eliminação recursiva de características (RFE), estabelecendo um limiar de concentração inibitória mínima (MIC < 50 $\mu\text{g/mL}$) como critério de corte. Entretanto, trata-se de uma prospecção ampla, não dedicada a alcaloides, e os modelos não foram desenvolvidos exclusivamente para prever atividade contra *F. oxysporum*.

2. JUSTIFICATIVA

A fusariose, causada por *Fusarium oxysporum*, representa um importante desafio fitossanitário para a agricultura brasileira, em razão de sua ampla distribuição geográfica, elevada persistência no solo e impacto significativo sobre culturas de alto valor econômico. As limitações dos métodos convencionais de controle, associadas a preocupações ambientais e à emergência de resistência, reforçam a necessidade de estratégias inovadoras para a identificação de compostos fungicidas mais eficazes e seguros.

Nesse contexto, a quimioinformática, por meio da modelagem de relações quantitativas entre estrutura e atividade (QSAR), destaca-se como uma abordagem racional e eficiente para a prospecção de novas substâncias bioativas. Assim que, os avanços na aplicação de métodos QSAR para o planejamento de PPPs e da reconhecida atividade antifúngica de alcaloides, não há modelos QSAR desenvolvidos especificamente para prever a atividade de alcaloides contra *Fusarium oxysporum*. O emprego de modelos QSAR possibilita a organização de dados experimentais, a seleção de descritores moleculares relevantes e o desenvolvimento de modelos preditivos robustos, reduzindo tempo e custos experimentais e aumentando a eficiência na triagem de candidatos promissores.

Assim, a presente dissertação justifica-se pela proposta de desenvolver modelos QSAR preditivos capazes de identificar alcaloides com potencial atividade fungicida contra *F. oxysporum*, contribuindo para o avanço metodológico da quimioinformática quanto para a busca de alternativas sustentáveis no manejo da fusariose.

3. OBJETIVOS

3.1 Objetivo Geral

Desenvolver modelo QSAR para a identificação racional de alcaloides com potencial de atividade fungicida para *Fusarium oxysporum* causador da doença fusariose.

3.2 Objetivos Específicos

- Estruturar banco de dados de alcaloides com atividade biológica contra *F. oxysporum*;
- Calcular e selecionar descritores moleculares para criação dos modelos QSAR;
- Desenvolver e validar modelos QSAR utilizando o algoritmo genético e *random forest* na previsão de modelos preditivos;
- Aplicar e analisar estruturas ativas definidas em triagem virtual para prospecção de novos candidatos bioativos contra *Fusarium oxysporum*.

4. METODOLOGIA

4.1 Softwares

Os cálculos foram realizados com o auxílio de um computador AMD Ryzen 5 2500U com 8GB de memória RAM, com sistema operacional *Microsoft Windows* 10. Os softwares utilizados foram:

- *AlvaDesk*, versão 2.0 (Alvascience, Itália), para calcular os descritores;
- *BIOVIA Discovery Studio* 2021 (Dassault Systemes, França), para leitura dos dados;
- *Knime Analytics Platform*, versão 5.2 (KNIME AG, Alemanha), para desenvolvimento e validação dos modelos;
- MOPAC (Molecular Orbital PACkage) versão 22.1.1, para calcular a conformação tridimensional;
- Open Babel, versão 3.1.1, (<http://openbabel.org>), para conversão de arquivos;
- RDKit – *Open-source Cheminformatics Software*, (<https://www.rdkit.org>) na preparação das estruturas;
- WEKA – *Waikato Environment for Knowledge Analysis*, (Universidade de Waikato, Nova Zelândia) pela extensão *KNIME Weka Data Mining Integration* (3.7) para desenvolver os modelos.

4.2 Métodos

4.2.1 Modelagem QSAR para *F. oxysporum*

4.2.1.1 Conjunto de dados inicial

Para realização dos estudos de modelagem QSAR, foi realizada a coleta de estruturas químicas a partir do ChEMBL (Zdrazil *et al.*, 2024) (<https://www.ebi.ac.uk/chembl/>) para *F. oxysporum*, caracterizada pelo ID: ChEMBL612648. O banco de dados selecionado para a atividade biológica de concentração inibitória mínima (MIC) continha um total de 997 estruturas testadas para o fungo.

4.2.1.2 Organização do Conjunto de Dados

Inicialmente, os dados obtidos foram carregados no *software* Knime Analytics Platform (Berthold *et al.*, 2008), um programa de código aberto que permite a fácil montagem visual e execução de *pipelines* de dados, com ferramentas de manipulação e visualização em forma de módulos ou nós. Foram filtradas inicialmente as estruturas cujas unidades estavam em $\mu\text{g.mL}^{-1}$, resultando em 767 estruturas. Em seguida, para selecionar estruturas similares aos alcaloides, foram filtradas apenas as estruturas que possuem átomos de C, H, N e O para desenvolvimento dos modelos, permanecendo 320 estruturas no banco de dados.

Posteriormente, realizou-se a padronização da unidade de medida das atividades MIC de $\mu\text{g.mL}^{-1}$ em unidades molares (mol/L). As unidades molares obtidas foram convertidas em unidades logarítmicas ($-\log\text{MIC}$ para pMIC), com o objetivo de reduzir a assimetria dos dados. Também foram realizadas: remoção de saís, remoção de estruturas duplicadas baseadas no ID e eliminação de *outliers* numéricos utilizando um multiplicador do intervalo interquartil de 1,5, além da estimativa de dados completos usando R_4 (Fourches; Muratov; Tropsha, 2010). Assim, foram selecionadas 299 estruturas.

As estruturas químicas remanescentes foram adicionadas hidrogênios, geradas suas coordenadas tridimensionais e otimizadas utilizando o campo de força UFF (Rappe *et al.*, 1992) com os pacotes do RDKit. Seguiu-se a otimização final das geometrias auxiliadas pelo MOPAC v. 22.1.1 (Moussa; Stewart, 2024) utilizando o método semiempírico PM7 com limite de norma de gradiente igual a 0,01 kcal/Å (Stewart, 2013). Foram obtidas as informações para cálculo de descritores moleculares: calor final de formação (kcal/mol), área e volume cosmo (em angstroms), norma do gradiente, potencial de ionização (eV), HOMO (eV) e LUMO (eV).

4.2.1.3 Cálculo dos descritores moleculares

Com as estruturas otimizadas e refinadas, utilizou-se o *software* AlvaDesc para cálculo de 5.762 descritores, cujas quantidades e descrições podem ser observadas no Quadro 2.

Quadro 2: Descritores moleculares e suas quantidades calculadas pelo Alvadesk para pMIC.

Categoria	Quantidade	Descrição
Índices Constitucionais	55	Refletem a composição elementar (contagem de átomos, ligações, peso molecular) sem considerar a conectividade.
Descritores de Anel	38	Informações sobre a presença, tamanho, tipo e fusão de sistemas cíclicos na molécula.
Índices Topológicos	90	Baseados na teoria dos grafos, descrevem o tamanho, forma e ramificação da molécula via matrizes de adjacência.
Contagem de Caminhadas e Trilhas	46	Representam a conectividade molecular através da sequência de átomos e ligações percorridas na estrutura.
Índices de Conectividade	37	Avaliam o grau de substituição e a natureza das ligações (ex: índices de Randic e Kier-Hall).
Índices de Informação	51	Quantificam a complexidade estrutural e o conteúdo de informação baseados na teoria da informação (entropia).
Baseados em Matriz 2D	608	Derivados de operações matemáticas em matrizes que representam a topologia bidimensional da molécula.
Autocorrelação 2D	225	Descrevem a distribuição de propriedades atômicas (massa, carga) ao longo da topologia molecular 2D.
Valores Eigen Burden (BCUT)	96	Valores próprios de matrizes de adjacência modificadas, usados para codificar conectividade e propriedades atômicas.
Tipo-P_VSA	96	Relacionam a área superficial de Van der Waals com propriedades físico-químicas específicas (<i>logP</i> , <i>molar refractivity</i>).

Índices ETA	40	Extended Topochemical Atom: descrevem o ambiente eletrônico e estérico de átomos individuais.
Adjacência de Arestas	324	Focam nas relações entre as ligações químicas (arestas) em vez dos átomos (nós).
Geométricos	39	Calculados a partir das coordenadas cartesianas 3D; incluem volume, forma e momentos de inércia.
Baseados em Matriz 3D	132	Derivados de matrizes que utilizam distâncias geométricas reais entre os átomos no espaço.
Autocorrelação 3D	80	Similar à 2D, mas considera a distância espacial real entre átomos no conformômero 3D.
RDF (Radial Distribution Function)	240	Descrevem a probabilidade de encontrar átomos a uma certa distância radial, independente da orientação.
3D-MoRSE	256	Codificam a estrutura 3D com base em transformadas matemáticas usadas em espalhamento de elétrons.
WHIM	387	Invariantes holísticos que capturam tamanho, forma, simetria e densidade de propriedades no espaço 3D.
Perfis Moleculares de Randic	41	Descrevem a forma molecular através de caminhos de diferentes comprimentos entre os átomos.
Contagem de Grupos Funcionais	153	Simple contagem de subestruturas específicas (ex: hidroxilas, aminas, anéis aromáticos).
Fragmentos Centrados de Átomo	115	Descrevem o ambiente químico imediato de cada tipo de átomo na estrutura.
Índices Tipo-Átomo Estado-E	346	Combinam informações eletrônicas e topológicas para descrever o estado eletrotológico dos átomos.
Farmacofóricos	165	Descrevem distâncias e arranjos entre grupos que interagem biologicamente (aceitadores/doadores de H, etc.).

Pares de Átomos 2D	1.595	Contagem de pares de tipos de átomos separados por uma distância topológica específica.
Pares de Átomos 3D	51	Contagem de pares de tipos de átomos separados por uma distância geométrica no espaço.
Propriedades Moleculares	29	Estimativas físico-químicas globais, como LogP, refratividade molar e solubilidade.
CATS 3D	300	Chemically Advanced Template Search: focado em farmacóforos tridimensionais para busca de ligantes.
WHALES	33	Baseados na distribuição de potencial eletrostático e forma da superfície molecular.
MDE (Molecular Distance Edge)	19	Índices de distância molecular entre tipos específicos de átomos.
Quiralidade	70	Medem a assimetria molecular e a presença de centros estereogênicos.
SASA (Solvent Accessible Surface Area)	31	Calculam a área da superfície da molécula que é acessível ao solvente.

Fonte: Mauri, 2020; Mauri; Bertola, 2022.

4.2.1.4 Tratamento dos descritores moleculares

Para esta etapa, foram removidos 25 descritores que apresentaram valores ausentes em pelo menos uma das estruturas e 202 descritores de contagem de números, segundo os autores Basak *et al.* (2006) e Rácz; Bajusz; Héberger (2019), descritores de contagem como número de determinados grupos funcionais ou tipos de átomos frequentemente apresentam distribuição assimétrica com muitos valores iguais a zero, o que pode comprometer a estabilidade estatística dos modelos e levar a superajuste (*overfitting*), e a ser altamente correlacionados entre si e com outros descritores, introduzindo redundância que dificulta a interpretação e a validação dos modelos. A seguir, os descritores passaram por um filtro de baixa variância, com limite superior de variação de 0,01, com o objetivo de garantir um bom modelo entre os dados, resultando em 3.029 descritores. Após o filtro, os descritores restantes foram analisados quanto à alta correlação de Pearson entre si, com utilização de *script* em Python para cálculo e exclusão dos descritores

com corte de 0,99, uma vez que valores entre 0,95 e 0,99 são recomendados para avaliação (Rácz; Bajusz; Héberger, 2019), permanecendo 1.693 descritores.

Em seguida, realizou-se a divisão das estruturas restantes a partir de valor de corte em ativas e inativas, testadas com as atividades dos antifúngicos clotrimazol, miconazol, anfotericina B e nistatina (Borrego-Muñoz; Ospina; Quiroga, 2021; Tzatzarakis *et al.*, 2001). O cálculo da atividade foi realizado a partir do banco inicial para desenvolvimento dos modelos, com seleção para cada antifúngico e seu identificador, contendo estruturas testadas com informações de peso molecular, notação SMILES e valores de MIC. Para cada antifúngico, os valores foram agrupados, transformados em mol/L e calculada a unidade logarítmica, igualmente à metodologia utilizada para os modelos. Foram obtidos os valores de pMIC: clotrimazol pMIC = 3,972; miconazol pMIC = 4,326; anfotericina B pMIC = 4,988; e nistatina pMIC = 4,858. Esses valores foram utilizados para teste na divisão de estruturas ativas (A) e inativas (B) como valor de corte. O melhor resultado para divisão foi obtido a partir do miconazol (ChEMBL91), com valor de pMIC = 4,326 e 8 estruturas testadas com valores de MIC presentes de 3,125 a 25 $\mu\text{g.mL}^{-1}$. As estruturas com valor acima de pMIC = 4,326 foram utilizadas na divisão "A" como ativas, e as com valor abaixo em "B" como inativas. Assim, obteve-se a divisão A = 57 estruturas ativas e B = 242 estruturas inativas. A partir dessa etapa, utilizou-se a classe AB para construção do modelo.

Utilizou-se, a seguir, a técnica de sobreamostragem minoritária sintética (SMOTE), desenvolvida por Chawla *et al.* (2002), com o objetivo de balancear e enriquecer o conjunto. Na definição da quantidade de vizinhos próximos, foram considerados em cada amostra de classe. O método cria linhas sintéticas entre um objeto de uma classe determinada e um de seus vizinhos mais próximos da mesma classe, selecionando um ponto ao longo da linha entre os dois objetos e determinando seus atributos ou valores do novo objeto baseado no ponto escolhido de forma aleatória. Foram utilizadas configurações de sobreamostragem para a classe A, minoritária na classe AB, considerando as 5 estruturas vizinhas mais próximas com a classe AB obtida na etapa anterior, adicionando 185 estruturas, resultando em 484 estruturas no banco de dados (Classe A = 242 e B = 242).

Foi proposta a divisão dos conjuntos das amostras em outros dois subconjuntos. Nesse particionamento, podem ser utilizadas as proporções 70% e 30% (Bosc *et al.*, 2019) ou 80% e 20% (Golbraikh *et al.*, 2016; Tropsha, 2010). O conjunto maior possui o objetivo de treinar o modelo (conjunto de treino ou treinamento) e o menor conjunto sendo responsável por testar o modelo (conjunto de teste). Diante do exposto, os conjuntos de dados foram testados em

particionamento 80%-20% e em formato 70%-30%. Onde o formato 70%-30% demonstrou melhor divisão pela estratégia estratificada, que realiza a preservação da distribuição dos valores na coluna do grupo para a construção do modelo. Sendo obtidas 338 estruturas para treino e 146 para teste.

Para a normalização dos dados, foi utilizado o método de mínimo-máximo de 0 a 1 para treino e teste, resultando em 1.692 descritores para seleção a seguir. Os dados são escalonados em um intervalo de 0 a 1, em que $\max(T)$ e $\min(T)$ são, respectivamente, valores máximos e mínimos da série T, x_i é uma observação da série T no instante i e x'_i é a observação na série T reescalada com o processo de normalização (Lima *et al.*, 2023), conforme a Equação 1 abaixo:

Equação 1: Normalização mínimo-máximo.

$$x'_i = \frac{x_i - \min(T)}{\max(T) - \min(T)} \quad (1)$$

4.2.1.5 Seleção dos descritores moleculares

O conjunto de dados foi utilizado como variável de entrada para seleção de subconjuntos dos descritores mais relevantes. Para a seleção de variáveis, foi realizada a técnica de regressão logística com regularização por operador de seleção e redução absoluta mínima (LASSO), integrando uma penalidade L1, mencionada por Tibshirani em 1996, com modelo linear e função de custo de mínimos quadrados. Assim, foram selecionados 27 descritores. A penalidade L1 provoca que um subconjunto dos pesos se torne zero, o que sugere que as características correspondentes podem ser descartadas (Goodfellow; Bengio; Courville, 2016).

Os 27 descritores selecionados pela regressão LASSO foram utilizados no Algoritmo Genético (AG) para a seleção final de descritores (Chen *et al.*, 2020). O AG compreende um método inicialmente desenvolvido por John H. Holland (Holland, 1992), sendo uma ferramenta eficiente muito utilizada para o desenho de fármacos e análise de dados, que demonstra resultados em predições de modelos robustos e precisos (Hemmateenejad *et al.*, 2002; Pourbasheer *et al.*, 2014).

Trata-se de um algoritmo inspirado na teoria da evolução, que considera princípios de seleção natural, cruzamento e mutação genética, compreendendo 5 etapas: (1) codificação das variáveis; (2) iniciação da população; (3) avaliação das respostas; (4) reproduções; e (5) mutações. As etapas de 3 a 5 são alternadas até que um critério de parada seja atingido. Este

critério pode se basear na falta de melhorias da resposta, em um número máximo de gerações ou no tempo total para elaboração permitido (Leardi *et al.*, 1992).

4.2.1.6 Construção dos modelos

Em estudos de Gramática (2020), recomenda-se que, para conjuntos de dados de 20 a 60 estruturas, sejam utilizadas proporções de estruturas e descritores sempre maiores que 10 a 15:1 no desenvolvimento de modelos. Os conjuntos de dados a partir de centenas devem partir da proporção de ao menos 50:1, aumentando conforme a quantidade de estruturas totais a serem consideradas na construção dos modelos. Assim, no AG foi utilizada a configuração de 5 a 10 descritores para melhor seleção, que auxiliem na explicação da variação da atividade biológica, utilizando o coeficiente Kappa de Cohen (κ) (Cohen, 1960) para medir a confiabilidade.

Os modelos foram gerados com auxílio do algoritmo *Random Forest* (RF), proposto por Breiman (2001), baseado na agregação de múltiplas árvores de decisão a partir da construção de amostras *bootstrap*. Cada árvore é treinada por uma amostra obtida do conjunto original e pela seleção aleatória de variáveis capazes de obter um resultado preditivo. Durante a construção de cada árvore e nó de decisão, um subconjunto aleatório de descritores é considerado na determinação da melhor divisão, impedindo que descritores com alta correlação tenham dominância na construção do modelo e promovendo diversidade entre as árvores (Breiman *et al.*, 2017).

4.2.2 Validação dos modelos

Os modelos desenvolvidos foram avaliados utilizando o coeficiente κ como métrica principal, além de matrizes de confusão, estatísticas gerais e de classe, acurácia global e área sobre a curva ROC.

4.2.2.1 Coeficiente Kappa de Cohen

O coeficiente Kappa de Cohen (κ) é um método inicialmente desenvolvido por Jacob Cohen em 1960. O autor menciona que os resultados de κ devem ser interpretados a partir dos valores contidos na Tabela 1, na avaliação do nível de concordância de dois parâmetros em duas categorias ou variáveis categóricas não ordenadas, partindo de 3 categorias ou mais (Li; Gao; Yu, 2023).

Tabela 1: Valores para interpretação do coeficiente Kappa de Cohen.

Valores de Kappa	Nível de concordância
0,00 – 0,20	Nenhuma
0,21 – 0,39	Mínima
0,40 – 0,59	Pequena ou fraca
0,60 – 0,79	Moderada
0,80 – 0,90	Forte
0,91 – 1,00	Quase perfeita

Fonte: Adaptado de Cohen, 1960; Landis & Koch, 1977.

Trata-se de um método utilizado para normalização da acurácia e redução da tendência, sendo uma técnica robusta que considera a ocorrência por classe (McHugh, 2012). O coeficiente Kappa entende-se como a proporção de desacordos esperados pelo acaso que não ocorrem, ou, de forma alternativa, a proporção da concordância após a concordância por acaso ser removida da consideração (Cohen, 1960). O coeficiente pode ser calculado a partir da Equação 2:

Equação 2: Cálculo do coeficiente Kappa de Cohen.

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (2)$$

$\text{Pr}(a)$ representa a concordância observada real e $\text{Pr}(e)$ a concordância aleatória esperada. A $\text{Pr}(e)$ pode ser calculada pela Equação 3 a seguir:

Equação 3: Cálculo da concordância aleatória esperada.

$$\text{Pr}(e) = \frac{\left(\frac{cm^1 \times rm^1}{n}\right) + \left(\frac{cm^2 \times rm^2}{n}\right)}{n} \quad (3)$$

Onde: cm^1 representa a margem da coluna 1, cm^2 representa a margem da coluna 2, rm^1 representa a margem da linha 1, rm^2 representa a margem da linha 2 e n representa o número de observações.

4.2.2.2 Matriz de Confusão

A matriz de confusão é uma ferramenta fundamental para avaliação de modelos de classificação binária, pois permite visualizar de forma estruturada os acertos e erros do classificador em relação aos valores reais (Kotu; Deshpande, 2019). No presente estudo, utilizou-

se a classe ativa (A) como classe positiva e a classe inativa (B) como classe negativa, conforme detalhado no Quadro 3.

Quadro 3: Matriz de confusão para uma classificação binária.

	Condição positiva	Condição negativa
Predição positiva	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Predição negativa	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Adaptado de Kotu; Deshpande, 2019.

Valores de Falso Positivo ocorrem quando o classificador estima o valor 0 para uma classe, mas ela verdadeiramente pertence à classe 1. Da mesma forma, Falso Negativo ocorre ao ser classificado como 1 e pertencer na verdade à classe 0 (Davis; Goadrich, 2006). Exemplificando a interpretação dos quadrantes ao contexto da prospecção de alcaloides podemos mencionar:

- Verdadeiro Positivo (VP): alcaloides classificados corretamente como ativos contra o *F. oxysporum*, representando os mais promissores para validação experimental.
- Verdadeiro Negativo (VN): alcaloides classificados corretamente como inativos. Embora não avancem para etapa de validação, sua identificação evita experimentos desnecessários.
- Falso Positivo (FP): alcaloides inativos que foram classificados de forma errada como ativos. Para triagem virtual, FPs podem representar compostos que serão levados para validação sem sucesso.
- Falso Negativo (FN): alcaloides ativos classificados incorretamente como inativos. Essa classificação representa a perda de compostos potencialmente bioativos que não serão considerados para validação experimental.

A partir dos valores da matriz de confusão, podem ser calculadas métricas derivadas que avaliam diferentes aspectos referentes ao desempenho do modelo (Kotu; Deshpande, 2019). Os principais parâmetros e suas definições serão mencionadas a seguir:

- Acurácia – habilidade do classificador em selecionar todos os casos necessários a serem selecionados e rejeitar todos os casos que precisam ser rejeitados (Kotu; Deshpande, 2019). Calculada a partir da Equação 4:

Equação 4: Cálculo da acurácia.

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} \quad (4)$$

- Especificidade – habilidade do classificador em rejeitar todas as classes que precisam ser rejeitadas. Em um classificador perfeito, rejeitaria todos os N reais e seu FP seria nulo (Kotu; Deshpande, 2019). Calculada a partir da Equação 5:

Equação 5: Cálculo da especificidade.

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad (5)$$

- Precisão – caracterizada como a proporção encontrada de casos realmente relevantes. Em um exemplo com 100 casos, se 80 são realmente relevantes, a precisão seria de 80/100 ou 80% (Kotu; Deshpande, 2019). Assim, 80 itens seriam VN e os 20 restantes FP. Calculada a partir da Equação 6:

Equação 6: Cálculo da precisão.

$$\text{Precisão} = \frac{VP}{VP+FP} \quad (6)$$

- *Recall* – definido a partir da proporção de casos relevantes encontrados entre os casos realmente relevantes. Em um exemplo, se apenas 50 de 100 casos relevantes foram encontrados, onde 50 casos relevantes foram encontrados e 50 foram perdidos, entrega-se um *recall* de 50% (Kotu; Deshpande, 2019). Calculado a partir da Equação 7:

Equação 7: Cálculo do *recall*.

$$\text{Recall} = \frac{VP}{VP+FN} \quad (7)$$

- Sensibilidade – habilidade do classificador em selecionar todos os casos que necessitam ser selecionados (Kotu; Deshpande, 2019). Calculada a partir da Equação 8:

Equação 8: Cálculo da sensibilidade.

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad (8)$$

Em contexto de triagem virtual de alcaloides, é importante visualizar o equilíbrio das métricas na matriz de confusão e otimizado pelo limitar de classificação (*threshold*), discutido na seção seguinte sobre a curva ROC. As métricas de sensibilidade, recall e precisão assumem relevância pois possuem objetivos em não perder compostos ativos, para minimizar a presença de FNs e evitar números excessivos de FPs que inviabilizam a triagem.

4.2.2.3 Área sob a curva ROC

A curva ROC (*Receiver Operating Characteristic Curve*) é uma representação gráfica que ilustra o desempenho de um classificador binário em todos os limiares de decisão (*thresholds*). Inicialmente desenvolvida por Green em 1966 para o campo de detecção de sinais de radar durante a segunda guerra mundial, posteriormente incorporada a psicologia (Tanner;

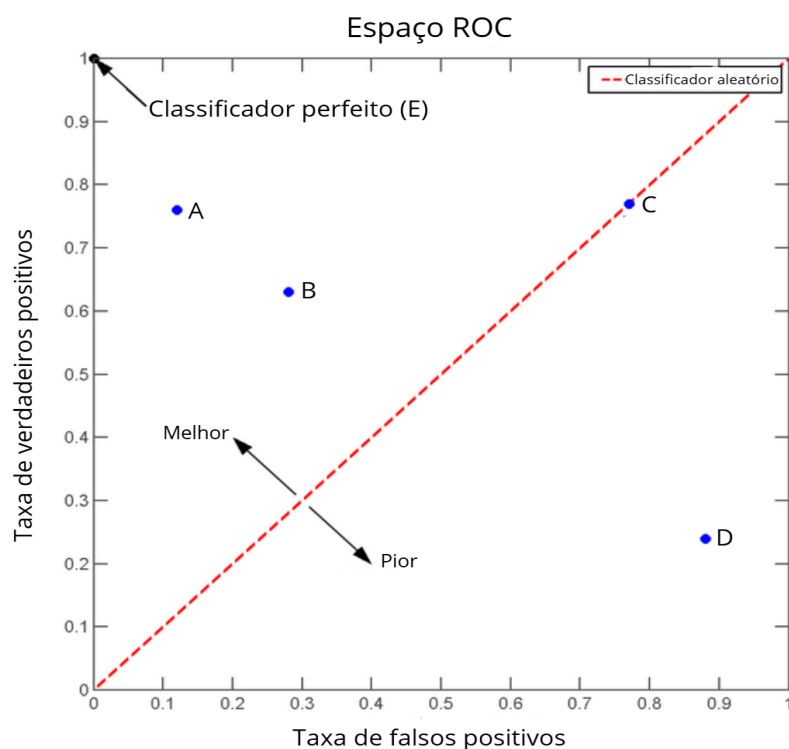
Swets, 1954), medicina na avaliação da performance de métodos diagnósticos (Lusted, 1971), a bioinformática e aprendizado de máquina (Sonogo; Kocsor; Pongor, 2008).

Para construção da curva são considerados valores de limiar, para cada valor temos uma probabilidade de corte utilizado para classificar um composto como ativo ou inativo, calculados a partir de dois parâmetros:

- Taxa de verdadeiros positivos (TVP) ou Sensibilidade.
- Taxa de falsos positivos (TFP) ou 1 - Especificidade.

A curva ROC pode ser construída ao plotar a TVP em eixo Y, contra a TFP em eixo X (Kotu; Deshpande, 2019), para todos os limiares possíveis, desde os mais restritivos como todos classificados como inativos, aos mais permissivos como todos classificados como ativos (Nahm, 2022). A Figura 4 ilustra este conceito com cinco classificadores (A, B, C, D e E), cada um produz um par de TVP e TFP que corresponde a um único ponto no espaço ROC. Como também, temos um classificador aleatório caracterizado pela linha tracejada em vermelho, onde são interpretados resultados acima da reta como melhores e abaixo da reta piores.

Figura 4: Gráfico ROC com cinco classificadores.



Fonte: Adaptado de Fawcett, 2006.

A área sob a curva ROC (*Area Under the Curve* - AUC) nas análises ROC pode ser interpretada como a probabilidade de o modelo atribuir uma pontuação mais alta a um composto ativo escolhido de forma aleatória do que a um composto inativo também escolhido aleatoriamente. Segundo Fawcett (2006), a AUC tem sua capacidade preditiva partindo da porção da área do quadrado, onde seu valor estará sempre entre 0 e 1,0, partindo de uma predição aleatória, é produzida uma linha diagonal de área 0,5, assim nenhum classificador realista deve ter AUC menor que 0,5. Os valores de AUC são interpretados conforme a Tabela 2:

Tabela 2: Interpretações dos valores de AUC.

Valores de AUC	Interpretação
$AUC < 0,5$	Pior que o aleatório
$AUC = 0,5$	Equivalente ao aleatório
$0,6 \leq AUC < 0,7$	Fraco
$0,7 \leq AUC < 0,8$	Aceitável
$0,8 \leq AUC < 0,9$	Bom
$0,9 \leq AUC < 1,0$	Excelente
$AUC = 1,0$	Perfeito

Fonte: Adaptado de Nahm, (2022); Rahnenführer *et al.*, (2023).

4.2.3 Triagem virtual

4.2.2.1 Banco de Alcaloides

Para realização da triagem virtual, utilizou-se um banco produzido por Paixão (2022) e estruturas oriundas do grupo de estudo, contendo 1.297 alcaloides, validadas na plataforma SciFinder (<https://scifinder-n.cas.org>). O banco possui diferentes classes de alcaloides, tais como: aporfínicos, indólicos, quinolínicos, oxindólicos e isoquinolínicos. O tratamento dos dados foi realizado da mesma forma que no conjunto de modelagem para a construção dos modelos QSAR, resultando em 980 alcaloides após o tratamento.

4.2.2.2 Domínio de aplicabilidade (DA) para o banco da triagem

O domínio de aplicabilidade é uma etapa fundamental para definir o espaço químico onde as previsões geradas serão confiáveis, com o objetivo de avaliar a precisão da confiabilidade do modelo a partir da avaliação das moléculas em sua relação ao "domínio" do modelo (Baskin *et al.*, 2010). Assim, as medições de similaridade são utilizadas para definir o domínio com base nas distâncias euclidianas entre as estruturas de treinamento e as estruturas de teste e/ou triagem

virtual. A distância de um composto de teste para seu vizinho no conjunto de treinamento é comparada ao limite de domínio de aplicabilidade predefinido (APD, do inglês *Applicability Domain*). Se a similaridade estiver além desse limite, a previsão é considerada não confiável (Zhang *et al.*, 2006). O APD é calculado a partir da Equação 9:

Equação 9: Cálculo do domínio de aplicabilidade.

$$APD = 'd' + Z\sigma \quad (9)$$

Para o cálculo de 'd' e σ tem-se: cálculo da média das distâncias euclidianas entre todos os pares de estruturas de treinamento. Em seguida, o conjunto de distâncias menores que a média é formulado. 'd' e σ são calculados como média e desvio padrão das distâncias. Z é um valor de corte empírico, com valor padrão = 0,5 (Afantitis *et al.*, 2011; Melagraki *et al.*, 2010).

Após o tratamento, os descritores moleculares foram calculados com auxílio do AlvaDesc e, em seguida, normalizados, utilizando os valores aplicados na normalização dos descritores do conjunto de treino. Posteriormente, foram separados os descritores nos modelos selecionados para os estudos de QSAR para a classe AB. Com a seleção dos descritores, os alcaloides foram filtrados pelo domínio de aplicabilidade com o nó Enalos inserido no Knime, sendo utilizados apenas os pertencentes ao DA. Em seguida, foram filtrados em relação aos alcaloides classificados como ativos na classe A e em relação ao valor de confiabilidade para a atividade.

5. RESULTADOS E DISCUSSÃO

5.1 Modelagem QSAR para *F. oxysporum*

5.1.1 Construção dos modelos

Como resultado, foi desenvolvido um modelo QSAR classificatório com abordagem *in silico*, utilizando técnicas de aprendizado de máquina envolvendo regressão L1, algoritmo genético e *random forest*, combinados a estratégias de balanceamento, seleção de descritores e critérios de avaliação, com a divisão das estruturas ativas em classe A e inativos em classe B.

A Tabela 3 apresenta os seis descritores moleculares selecionados pelo modelo QSAR para a classe AB e suas respectivas importâncias relativas. Segundo Bajželj; Novič; Drgan. (2025), a importância relativa é considerada uma métrica fundamental para a quantificação da contribuição de cada descritor molecular na capacidade preditiva do modelo, indicando as características estruturais que realizam maior influência na discriminação entre os compostos ativos e inativos.

Tabela 3: Importância relativa dos descritores selecionados para classe AB.

Descritores moleculares	Importância Relativa
minssCH2	2,382
AROM	1,860
MaxsssN	1,198
VE3sign_G	1,151
SaasN	1,081
B06[N-O]	0,474

Fonte: Autor.

Em seguida, é realizada uma análise aprofundada de cada um dos descritores selecionados, interpretando seu significado químico e sua contribuição para a potencial atividade antifúngica contra *Fusarium oxysporum*.

5.1.1.2 Descritor minssCH2

O descritor minssCH2 representa o valor mínimo do átomo do Estado-E para o fragmento “-CH2-” em duas dimensões (2D). Descritores Estado-E (Eletropológicos) combinam informações sobre o estado eletrônico do átomo e sua conectividade eletropológica na molécula (Ibrahim *et al.*, 2020). Os parâmetros para esse tipo de átomo fornecem informações de densidade e distribuição de elétrons, disposição espacial de átomos eletronegativos e a capacidade de doar ou aceitar elétrons em uma reação química em uma molécula (Grisoni *et al.*, 2018). A alta importância do descritor indica que a presença mínima de fragmentos “-CH2-” são cruciais para a atividade.

Em estudo de Sandhu *et al.* (2022), o descritor foi identificado com coeficiente positivo no desenvolvimento de modelo QSAR preditivo para inibidores da acetilcolinesterase, uma enzima que realiza degradação da acetilcolina, um potencial fármaco para o tratamento da doença de Alzheimer. Em contexto antifúngico, a presença e o ambiente eletrônico de fragmentos “-CH2-” podem influenciar na flexibilidade e na capacidade molecular em realizar interações hidrofóbicas em alvos proteicos do fungo, como enzimas de parede celular ou membranas. Sua alta importância relativa de valor = 2,382 sugere que pequenas variações nesses grupos podem resultar em grandes diferenças na atividade inibitória, como ao encaixe molecular no sítio ativo (Anyubaga *et al.*, 2024).

5.1.1.3 Descritor AROM

O descritor AROM refere-se ao índice de aromaticidade ou à contagem de anéis aromáticos na molécula em índices de grau de deslocalização (Mauri, 2020). Compostos aromáticos são frequentemente encontrados em moléculas bioativas, pois seus anéis planares podem participar de interações de empilhamento π - π com aminoácidos aromáticos no sítio ativo de enzimas (Meunier *et al.*, 2022). Estudos de Meunier *et al.*, (2022) demonstram que muitos fármacos de pequenas moléculas interagem com sítios ativos de proteínas através de interações π - π entre o ligante e aminoácidos como triptofano, tirosina ou fenilalanina. Interações envolvendo sistemas π são reconhecidamente cruciais para eventos biológicos como o reconhecimento proteína-ligante e a estabilidade estrutural de proteínas (Meunier *et al.*, 2022). Sua alta importância de aromaticidade está alinhada ao estudo de Bernal e Coy Barrera (2019), com o desenvolvimento de modelo 3D-QSAR para alcaloides quinolizidínicos contra *F. oxysporum*, e realizaram a identificação da importância de anéis heterocromáticos, em exemplo a porção 2-piridona para atividade antifúngica. O valor de importância relativa em 1,860 para

AROM sugere que a planaridade e a capacidade de formar interações do tipo π são um fator determinante para as atividades dos compostos em estudo.

5.1.1.4 Descritor MaxsssN

O descritor MaxsssN representa o valor máximo do átomo do Estado-E para o nitrogênio ligado a três outros átomos por ligação simples (Mauri, 2020), como exemplo temos esses tipos de ligação em aminas e em diversos alcaloides. A alta relevância do descritor menciona o papel fundamental de átomos de nitrogênio na atividade antifúngica. Estudos de Vazhev *et al.* (2021) sobre a atividade de compostos orgânicos contra *F. oxysporum*, a presença de grupos funcionais contendo nitrogênio, como aminas, foi associada indiretamente a atividade, uma vez que a ausência ou modificação desses grupos pode levar a perda de eficácia. Sua alta importância sugere que o ambiente eletrônico dos nitrogenados são otimizados no conjunto de compostos ativos, possivelmente na atuação como aceptores ou doadores de ligações de hidrogênio.

5.1.1.5 Descritor VE3sign_G

VE3sign_G é um descritor topológico 2D, relacionado a soma dos coeficientes logarítmicos do último autovetor a partir da matriz geométrica/distância, inserido em descritores baseados na matriz 3D (Mauri, 2020). Uma matriz de distância reflete na distância entre os pares de átomos presentes em uma molécula de cada elemento, e seus autovetores presentes podem ser analisados em forma e tamanho, para revelar características estruturais e de simetria (Chebotaev *et al.*, 2024).

Sua importância relativa em 1,151 indica que a forma tridimensional da molécula e a distribuição de suas propriedades eletrônicas no espaço são fatores determinantes para a atividade antifúngica contra *F. oxysporum*. Descritores desta natureza são particularmente úteis em modelagem QSAR por capturarem características moleculares que não podem ser adequadamente representadas por descritores puramente bidimensionais ou topológicos (Mauri, 2020). Conforme destacado por Mauri (2020), a inclusão de descritores 3D como o VE3sign_G permite uma descrição mais completa e realista da estrutura molecular. Em contexto para inibição de *F. oxysporum*, a relevância deste descritor sugere que a interação com o alvo biológico depende de um encaixe espacial preciso entre o ligante e o sítio ativo da enzima fúngica.

5.1.1.6 Descritor SaasN

Semelhante ao MaxsssN, o descritor SaasN descreve o somatório dos estados eletrotológicos aos átomos de nitrogênio na molécula, ligados à presença de nitrogênio e sua contribuição às propriedades globais da molécula (Mauri, 2020). Portanto, SaasN representa a

contribuição total de nitrogênios que fazem parte de um anel aromático (Ex: nitrogênio em piridina, pirrol, ou em anéis fusionados de alcaloides como a citisina).

Este descritor, juntamente com MaxsssN e AROM, reforça a importância dos nitrogênios em ambientes alifáticos e aromáticos. A alta importância de SaasN em 1,081 corrobora fortemente os achados de Bernal e Coy-Barrera (2019), que previram que alcaloides derivados de citisina (que contêm um anel de piridona com um nitrogênio aromático) e com porção 2-piridona apresentariam a maior atividade antifúngica contra *F. oxysporum*. A soma dos estados eletrônicos desses nitrogênios pelo descritor SaasN, sugere estar presente como um determinante para a bioatividade, possivelmente por participarem de interações específicas com o alvo, como ligações de hidrogênio ou interações π cátion.

5.1.1.7 Descritor B06[N-O]

O descritor B06[N-O] contabiliza a presença ou ausência de pares de átomos e suas ocorrências de todos os pares de nitrogênio e oxigênio (N e O) separados por uma distância topológica de 6 ligações na molécula (Carhart; Smith; Venkataraghavan, 1985). Embora sua importância relativa seja a menor da lista (0,474), sua presença no modelo final indica que a distância específica entre átomos de nitrogênio e oxigênio contribui para a discriminação entre compostos ativos e inativos. Estes pares de átomos são características de grupos funcionais como amidas, nitro, ou ésteres, e a distância específica de 6 ligações pode representar o espaçamento ideal entre um grupo doador e um aceptor de ligação de hidrogênio para interagir com um sítio bioativo no fungo.

Estudos de modelagem demonstram que a inclusão de tais descritores de pares permite capturar características conformacionais sutis que não seriam adequadamente representadas por descritores globais (Sliwoski *et al.*, 2016; Wildman & Crippen, 2002). Conforme observado por Sliwoski *et al.* (2016), a limitação da distância máxima de pares de átomos (neste caso 6 ligações) pode aumentar o desempenho do modelo, possivelmente pois distâncias maiores introduzem maior flexibilidade conformacional e ruído. A inclusão deste descritor, mesmo com menor peso, contribui para o refinamento da precisão preditiva do modelo, capturando interações específicas entre grupos funcionais polares que são relevantes para o reconhecimento molecular pelo alvo fúngico.

5.2 Validação dos modelos

A validação é uma etapa crucial no desenvolvimento de modelos QSAR, pois garante que o modelo seja robusto, preditivo e não tenha sofrido superajuste (*overfitting*) aos dados de

treinamento (Tropsha, 2010; Gramatica, 2020). Neste estudo, foram avaliados parâmetros estatísticos de validação interna, cruzada e externa, matriz de confusão com sensibilidade, especificidade, precisão, coeficiente κ , análise de curva ROC e distribuição das probabilidades das classes AB.

5.2.1 Validação Interna

Inicialmente, serão avaliados os parâmetros de validação interna, avaliando seu desempenho na classificação de compostos ativos (classe A) e inativos (classe B) contra *Fusarium oxysporum*. Temos na Tabela 4 a matriz de confusão para o conjunto de validação interna e as estatísticas derivadas sumarizadas.

Tabela 4: Parâmetros avaliados na validação interna do modelo QSAR desenvolvido.

Validação Interna				
Matriz de Confusão				
	A (Real)	B (Real)		
A (Predito)	169	11		
B (Predito)	0	158		
Estatísticas de Classe				
Classe	Recall	Precisão	Sensibilidade	Especificidade
A	100.00%	93.89%	100.00%	93.49%
B	93.49%	100.00%	93.49%	100.00%
Estatísticas Gerais				
Precisão geral	Erro geral	Kappa de Cohen (κ)	Classificados corretamente	Classificados incorretamente
96.75%	3.25%	0.935	327	11

Fonte: Autor.

A matriz de confusão revela um desempenho satisfatório do modelo na classificação dos compostos. Dos 338 compostos avaliados, 327 (96,75%) foram corretamente classificados, enquanto apenas 11 (3,25%) apresentaram erro de predição. Em relação aos Falsos Positivos, onze compostos inativos (classe B real) foram erroneamente classificados como ativos (classe A predita). No contexto da triagem virtual, falsos positivos representam compostos que seriam selecionados para validação experimental sem apresentar atividade real. Entretanto, uma taxa de FP de apenas 6,51% (11 de 169 inativos) pode ser considerada muito baixa e perfeitamente

aceitável em estudos de prospecção, especialmente quando comparada a outros modelos QSAR publicados (Neves *et al.*, 2016; Mukherjee; Kumar; Roy, 2022). Para os Falsos Negativos, não apresentaram nenhum falso negativo, ou seja, todos os compostos ativos foram corretamente identificados pelo modelo. Sendo um resultado particularmente relevante para a prospecção de alcaloides, pois a classe de maior interesse (ativos) foi integralmente preservada. Assim, a ausência de FNs confere ao modelo uma alta confiabilidade para triagem virtual.

Para as métricas de classe temos a classe AB, nela foram avaliadas:

Classe Ativa (A):

- Recall/Sensibilidade = 100,00%: O modelo identificou todos os compostos ativos presentes no conjunto. Este valor máximo é consistente com a ausência de falsos negativos e indica que o modelo é altamente sensível para detectar compostos com potencial antifúngico.
- Precisão = 93,89%: Dos compostos classificados como ativos, 93,89% eram realmente ativos. Os 6,11% restantes (11 compostos) eram falsos positivos. Este valor de precisão é considerado excelente em modelagem QSAR, onde valores acima de 80% são frequentemente reportados como satisfatórios (Gramatica, 2020).

Classe Inativa (B):

- Especificidade = 100,00%: O modelo classificou corretamente todos os compostos inativos como pertencentes à classe B. Isto significa que nenhum composto ativo foi erroneamente classificado como inativo (FN = 0), o que é consistente com a análise anterior.
- Recall = 93,49%: Dos compostos inativos reais, 93,49% foram corretamente identificados. Os 6,51% restantes (11 compostos) foram os falsos positivos já discutidos.
- Precisão = 100,00%: Todos os compostos classificados como inativos eram de fato inativos, o que significa que não houveram falsos negativos.

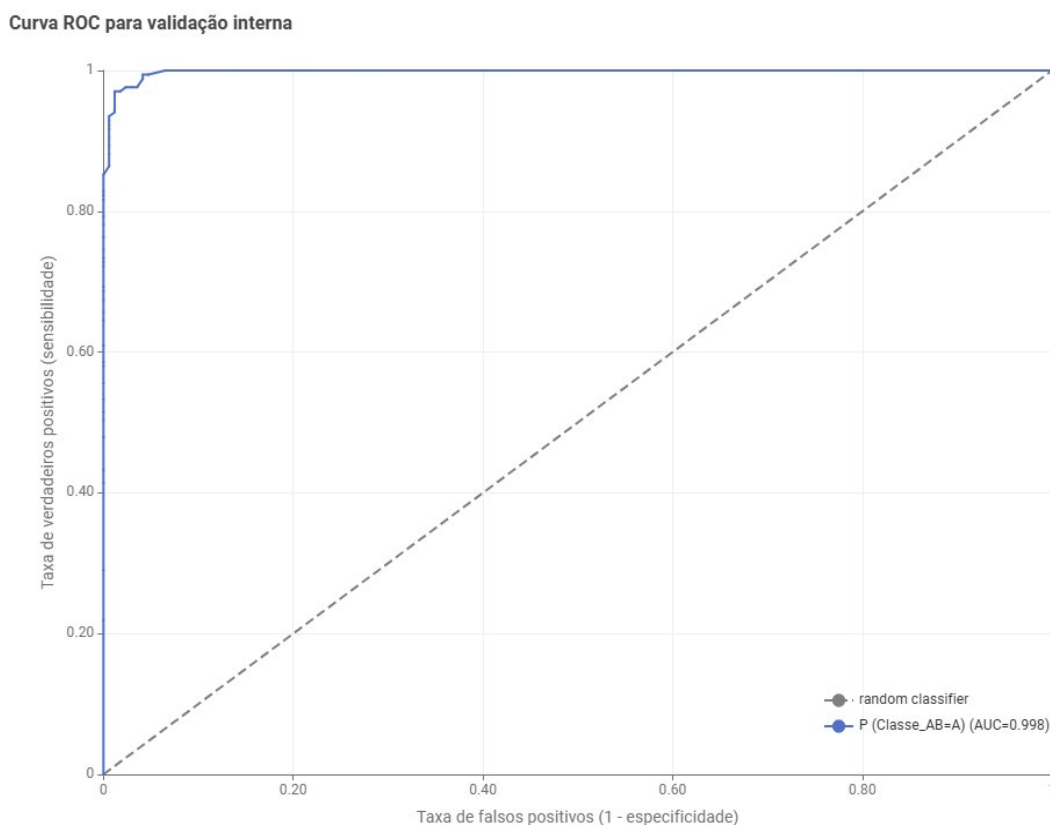
Para as estatísticas gerais, temos o coeficiente Kappa de Cohen (κ), uma métrica robusta que mede a concordância entre as predições do modelo e os valores reais, descontando a concordância esperada pelo acaso (Cohen, 1960). O valor obtido foi $\kappa = 0,935$, segundo a escala de interpretação proposta por Cohen (1960) e Landis & Koch (1977), é classificado como "concordância quase perfeita" conforme a Tabela 1. O valor de κ indica que o modelo é

extremamente confiável e que sua acurácia (96,75%) não se deve ao acaso. Este valor é comparável aos reportados em estudos recentes de QSAR para antifúngicos. Por exemplo:

- Neves *et al.* (2016) desenvolveram modelos QSAR para inibidores da enzima SmTGR de *Schistosoma mansoni* e reportaram valores de κ entre 0,70 e 0,85 para seus melhores modelos.
- Mukherjee, Kumar e Roy (2022), em um estudo de QSAR para predição de toxicidade de pesticidas em abelhas, obtiveram modelos com κ variando de 0,75 a 0,88.

Em seguida, temos a análise da curva ROC (Receiver Operating Characteristic) para o modelo na validação interna é apresentada na Figura 5, juntamente com a área sob a curva (AUC). Onde temos no Eixo X a Taxa de Falsos Positivos (1 - Especificidade), no Eixo Y a Taxa de Verdadeiros Positivos (Sensibilidade), na linha azul o modelo desenvolvido (AUC = 0,998) e a linha cinza um classificador aleatório (AUC = 0,5).

Figura 5: Curva ROC da validação interna para classe A.



Fonte: Autor.

A curva ROC do modelo (linha azul) apresenta um comportamento próximo do ideal, passando muito perto do canto superior esquerdo do gráfico, que representa o ponto de desempenho perfeito (100% de sensibilidade e 0% de falsos positivos). A curva afasta-se

significativamente da diagonal (classificador aleatório), indicando excelente poder discriminatório conforme Tabela 2. A área sob a curva ROC (AUC) foi calculada em 0,998. Este valor pode ser interpretado como a probabilidade de que o modelo atribua uma pontuação mais alta a um composto ativo escolhido aleatoriamente do que a um composto inativo também escolhido aleatoriamente (Nahm, 2022). Assim, o modelo situa-se na categoria de "excelente poder discriminatório", muito próximo do classificador perfeito (AUC = 1,0), sendo capaz de separar quase que perfeitamente a classe ativa e inativa.

Em comparação com a literatura, valores de AUC acima de 0,9 são considerados excelentes em modelagem QSAR. Temos alguns exemplos na literatura:

- Gramatica (2020), em uma revisão sobre boas práticas em QSAR, sugere que modelos com $AUC > 0,8$ são adequados para uso preditivo, enquanto $AUC > 0,9$ indica excelente desempenho.
- Neves *et al.* (2016) reportaram AUC entre 0,85 e 0,95 para seus modelos de inibidores da SmTGR.
- Chen *et al.* (2025), em um estudo de machine learning para inibidores de *Fusarium* spp., obtiveram AUC variando de 0,87 a 0,93 para diferentes algoritmos.
- Mukherjee, Kumar e Roy (2022) reportaram AUC entre 0,82 e 0,91 para seus modelos de toxicidade em abelhas.

5.2.2 Validação Cruzada

A validação cruzada é uma etapa fundamental para avaliar a robustez e a capacidade de generalização de um modelo QSAR, pois testa o desempenho do modelo em subconjuntos de dados não utilizados durante o treinamento, simulando seu comportamento diante de novos compostos (Tropsha, 2010; Gramatica, 2020). Neste estudo, foi empregada a validação cruzada estratificada com divisão 70%/30%, conforme descrito na metodologia. Os resultados obtidos para matriz de confusão e estatísticas estão apresentados a seguir na Tabela 5. A matriz de confusão da validação cruzada revela um desempenho ainda muito satisfatório, com 296 dos 338 compostos (87,57%) corretamente classificados.

Tabela 5: Parâmetros avaliados na validação cruzada do modelo QSAR desenvolvido.

Validação Cruzada				
Matriz de Confusão				
	A (Real)	B (Real)		
A (Predito)	157	30		
B (Predito)	12	139		
Estatísticas de Classe				
Classe	Recall	Precisão	Sensibilidade	Especificidade
A	92.90%	83.96%	92.90%	82.25%
B	82.25%	92.05%	82.25%	92.90%
Estatísticas Gerais				
Precisão geral	Erro geral	Kappa de Cohen (κ)	Classificados corretamente	Classificados incorretamente
87.57%	12.43%	0.751	296	42

Fonte: Autor.

Observa-se uma pequena queda em relação à validação interna (96,75%), o que era esperado, pois a validação cruzada avalia o modelo em dados não vistos durante o treinamento, fornecendo uma estimativa mais realista de seu desempenho futuro (Hawkins, 2004). Em relação aos Falsos Positivos e Falsos Negativos observados, podemos analisar:

- Falsos Positivos: Trinta compostos inativos (classe B real) foram erroneamente classificados como ativos (classe A predita). Isso representa 17,75% dos inativos reais (30 de 169). Embora superior aos 6,51% observados na validação interna, esta taxa ainda é considerada aceitável para um modelo de triagem virtual, especialmente considerando que a prioridade é a captura de ativos verdadeiros (Baumann & Baumann, 2014).
- Falsos Negativos: Doze compostos ativos (classe A real) foram incorretamente classificados como inativos. Este é um ponto de atenção, pois representa uma perda de 7,10% dos compostos ativos (12 de 169) que não seriam selecionados para validação experimental.

Para as métricas de classe temos a classe AB, foram avaliadas:

Classe Ativa (A):

- Recall/Sensibilidade = 92,90%: O modelo identificou corretamente 92,90% dos compostos ativos. Este valor é elevado e indica que o modelo mantém boa sensibilidade mesmo em dados não vistos, embora inferior aos 100% da validação interna.
- Precisão = 83,96%: Dos compostos classificados como ativos, 83,96% eram realmente ativos. Os 16,04% restantes (30 compostos) eram falsos positivos. Este valor ainda é bom, considerando que em modelos QSAR valores de precisão acima de 80% são frequentemente reportados como satisfatórios (Gramatica, 2020).

Classe Inativa (B):

- Especificidade = 92,90%: O modelo classificou corretamente 92,90% dos compostos inativos, indicando boa capacidade de rejeitar compostos inativos.
- Recall = 82,25%: Dos compostos inativos reais, 82,25% foram corretamente identificados. Os 17,75% restantes (30 compostos) foram os falsos positivos já discutidos.
- Precisão = 92,05%: Dos compostos classificados como inativos, 92,05% eram de fato inativos. Os 7,95% restantes (12 compostos) eram falsos negativos.

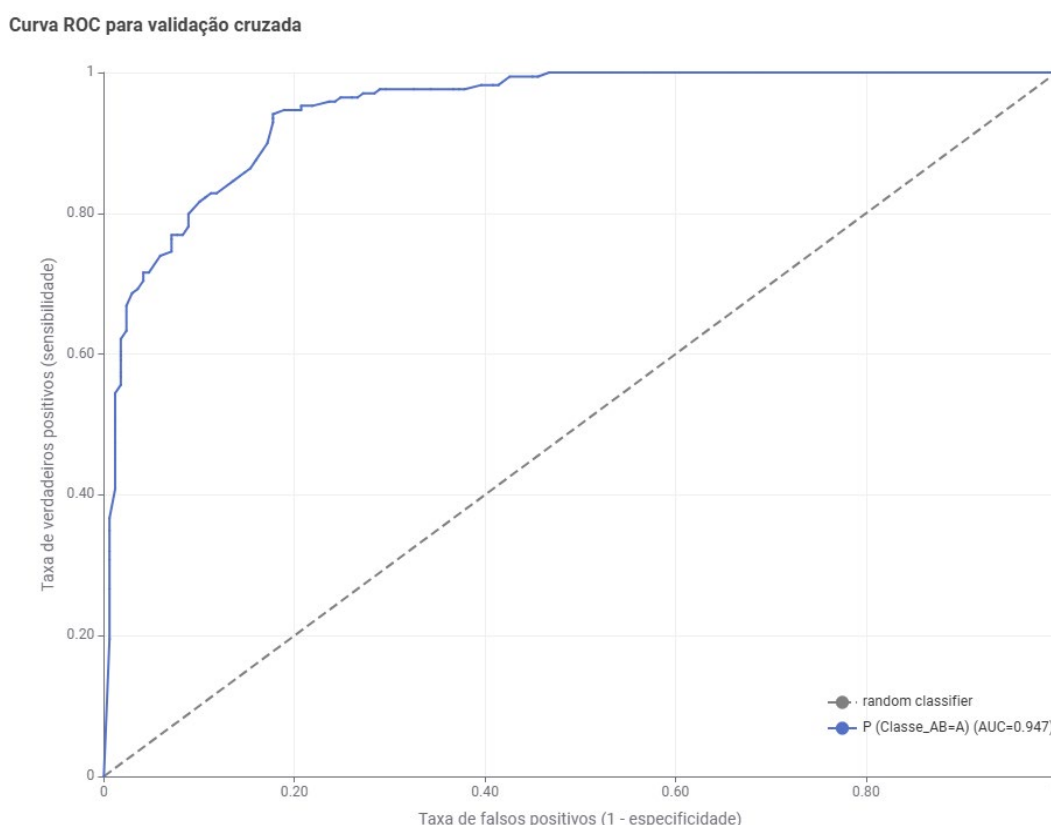
Em relação ao coeficiente Kappa de Cohen para a validação cruzada foi $\kappa = 0,751$. Segundo a escala de Landis & Koch (1977), este valor é classificado como "concordância substancial" (0,61 – 0,80). E indica que a concordância entre as predições do modelo e os valores reais é substancial e significativamente superior ao acaso. Sendo um valor consistente com modelos QSAR robustos reportados na literatura:

- Neves *et al.* (2016) obtiveram κ entre 0,70 e 0,85 em seus modelos para inibidores da SmTGR.
- Mukherjee, Kumar e Roy (2022) reportaram κ variando de 0,75 a 0,88 para modelos de toxicidade em abelhas.
- Borrego-Muñoz, Ospina e Quiroga (2021), em estudo com *F. oxysporum*, não reportaram κ , mas a acurácia de seus modelos (85-90%) sugere valores de κ provavelmente na faixa de 0,70-0,80.

O valor obtido de 0,751 situa-se, portanto, dentro da faixa esperada para modelos preditivos confiáveis e confirma a robustez do modelo desenvolvido.

A curva ROC para o modelo na validação cruzada é apresentada na Figura 6, juntamente com a área sob a curva (AUC). Temos o classificador aleatório de $AUC = 0,5$ em linha cinza e em linha azul para o modelo desenvolvido na validação cruzada.

Figura 6: Curva ROC da validação cruzada para classe A.



Fonte: Autor.

A curva ROC na validação cruzada (linha azul) mantém um bom comportamento, afastando-se de forma considerável da diagonal (classificador aleatório) e se aproximando do canto superior esquerdo. A área sobre a curva ROC (AUC) foi calculada em 0,947. Conforme Tabela 2 o modelo mantém a classificação de “excelente poder discriminatório”, confirmando sua capacidade de separar as classes ativa e inativa mesmo em dados não utilizado no treinamento. A redução em relação a AUC da validação interna (0,998) é esperada e reflete a maior dificuldade à predição com novos dados. Em comparação a estudos da literatura, os valores de AUC na faixa de 0,940-0,950 são considerados excelentes em modelagem QSAR:

- Gramatica (2020) sugere que modelos com $AUC > 0,9$ são adequados para uso preditivo e indicam alta capacidade discriminatória.
- Chen *et al.* (2025), em estudo de machine learning para inibidores de *Fusarium* spp., obtiveram AUC entre 0,87 e 0,93 para diferentes algoritmos, com o melhor modelo atingindo $AUC = 0,93$.
- Mukherjee, Kumar e Roy (2022) reportaram AUC máxima de 0,91 para seus modelos de toxicidade.
- Neves *et al.* (2016) obtiveram AUC entre 0,85 e 0,95, com os melhores modelos atingindo 0,94.

A redução observada nos indicadores de desempenho da validação interna para a validação cruzada é um fenômeno esperado (Tropsha, 2010). A validação interna tende a superestimar o desempenho do modelo, pois avalia o modelo nos mesmos dados utilizados para seu treinamento. A validação cruzada, ao testar o modelo em subconjuntos não vistos, fornece uma estimativa mais realista e conservadora do desempenho em novos compostos. Pelo modelo ainda apresentar métricas elevadas na validação cruzada (acurácia $> 87\%$, $\kappa > 0,75$, $AUC > 0,94$) indica que não houve superajuste significativo e que o modelo é robusto e generalizável. A diferença de 9,18% na acurácia entre as duas validações está dentro do esperado para modelos QSAR bem construídos (Gramatica, 2020).

5.2.3 Validação Externa

A validação externa é extremamente importante para avaliação de modelos QSAR, pois testa o modelo em um conjunto de dados que não foi utilizado em nenhuma etapa do processo de construção, seleção de descritores ou otimização de parâmetros (OECD, 2014; Tropsha, 2010). Diferente da validação cruzada, que ainda utiliza partições do conjunto utilizado no modelo, a validação externa utiliza compostos independentes, que simulam condições reais de aplicação do modelo a novas moléculas (Gramatica, 2020). Nesse estudo, o modelo foi testado em um conjunto de dados externo contendo 146 compostos, com seus resultados apresentados a seguir na Tabela 6, com matriz de confusão e estatísticas de classe e gerais.

A matriz de confusão revela um desempenho satisfatório, com 127 dos 146 compostos (86,99%) classificados corretamente. Como também, a acurácia manteve-se próxima a obtida na validação cruzada de 87,57%, indicando excelente estabilidade e generalização.

Tabela 6: Parâmetros avaliados para validação externa do modelo QSAR desenvolvido.

Validação Externa				
Matriz de Confusão				
	A (Real)	B (Real)		
A (Predito)	65	11		
B (Predito)	8	62		
Estatísticas de Classe				
Classe	Recall	Precisão	Sensibilidade	Especificidade
A	89.04%	85.53%	89.04%	84.93%
B	84.93%	88.57%	84.93%	89.04%
Estatísticas Gerais				
Precisão geral	Erro geral	Kappa de Cohen (κ)	Classificados corretamente	Classificados incorretamente
86.99%	13.01%	0.740	127	19

Fonte: Autor.

Em análise dos Falsos Positivos e Falsos Negativos obtidos, temos:

- Falsos Positivos: Onze compostos inativos (classe B real) foram erroneamente classificados como ativos (classe A predita). Isso representa 15,07% dos inativos reais (11 de 73). Esta taxa é ligeiramente inferior à observada na validação cruzada (17,75%) e ainda dentro de limites aceitáveis para triagem virtual (Baumann & Baumann, 2014).
- Falsos Negativos: Oito compostos ativos (classe A real) foram classificados de forma incorreta como inativos. Representando 10,96% dos ativos reais (8 de 73). Embora ligeiramente superior aos 7,10% da validação cruzada, ainda é uma taxa baixa e aceitável, especialmente considerando a independência do conjunto de teste.

Para as métricas de classe temos a classe AB, foram avaliadas:

Classe Ativa (A):

- **Recall/Sensibilidade:** O modelo identificou corretamente 89,04% dos compostos ativos no conjunto externo. Este valor é elevado e muito próximo ao obtido na validação cruzada (92,90%), confirmando a capacidade do modelo em capturar a maioria dos compostos potencialmente bioativos.
- **Precisão:** Dos compostos classificados como ativos, 85,53% eram realmente ativos. Os 14,47% restantes (11 compostos) eram falsos positivos. Este valor é ligeiramente superior ao da validação cruzada (83,96%), indicando boa confiabilidade nas predições positivas.

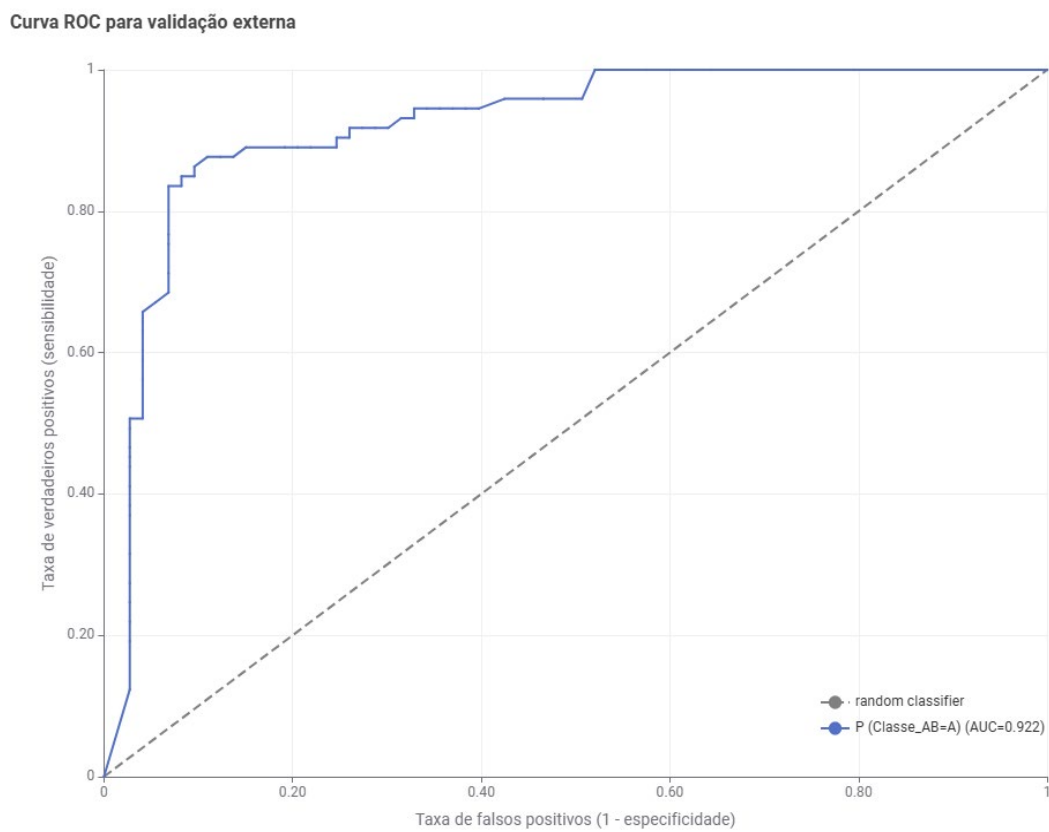
Classe Inativa (B):

- **Especificidade:** O modelo classificou corretamente 89,04% dos compostos inativos, valor muito próximo aos 92,90% da validação cruzada.
- **Recall:** Dos compostos inativos reais, 84,93% foram corretamente identificados, ligeiramente superior aos 82,25% da validação cruzada.
- **Precisão:** Dos compostos classificados como inativos, 88,57% eram de fato inativos. Os 11,43% restantes (8 compostos) eram falsos negativos.

O coeficiente Kappa de Cohen para a validação externa foi $\kappa = 0,740$. Segundo a Tabela 2 temos um valor classificado como "concordância substancial" (0,61 – 0,80). O valor de κ indica que a concordância entre as predições do modelo e os valores reais no conjunto externo é substancial e significativamente superior ao acaso e extremamente próximo ao obtido na validação cruzada (0,751), demonstrando excelente consistência e reprodutibilidade do modelo.

Para curva ROC da validação externa, na Figura 7 podemos observar que a linha azul mantém um comportamento satisfatório, afastando significativamente da diagonal e posicionando-se no quadrante superior esquerdo, com área sobre a curva ROC (AUC) calculada em 0,922. Com essa AUC o modelo mantém a classificação de “excelente poder classificatório”, mesmo em conjunto externo, confirmando sua capacidade de separar as classes ativa e inativa em dados novos. A pequena redução observada em relação a AUC da validação cruzada em 0,947 é esperada e reflete na maior dificuldade à predição de compostos independentes

Figura 7: Curva ROC da validação externa para classe A.



Fonte: Autor.

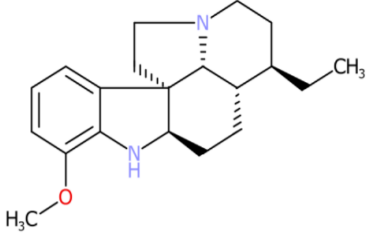
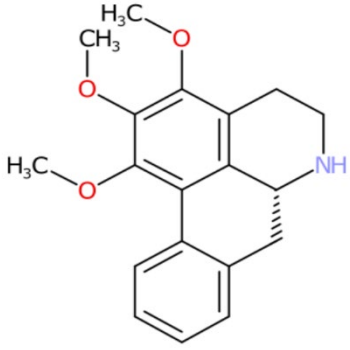
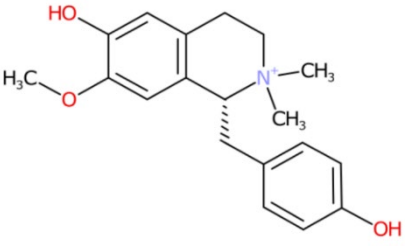
De forma integrada, os resultados das três etapas de validação indicam que o modelo QSAR desenvolvido é estatisticamente robusto, estável e preditivamente confiável, atendendo aos critérios recomendados para modelos classificatórios. A manutenção de elevados valores de AUC, associada a coeficientes Kappa satisfatórios e boas métricas de sensibilidade e especificidade, sugere que os descritores selecionados capturam informações estruturais relevantes para a discriminação entre as classes estudadas.

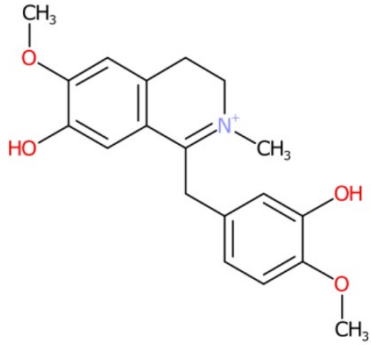
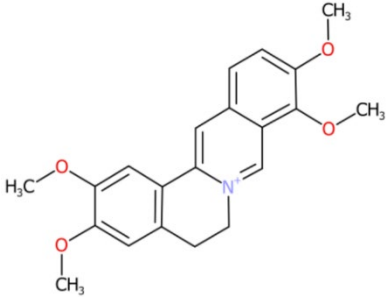
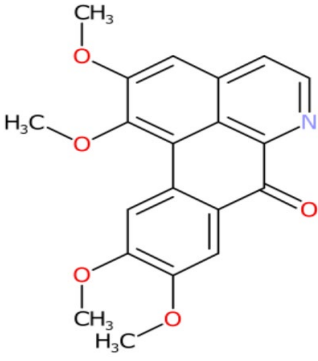
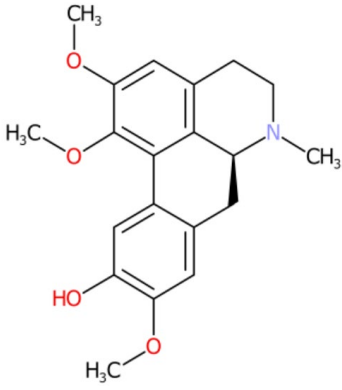
5.3 Triagem Virtual

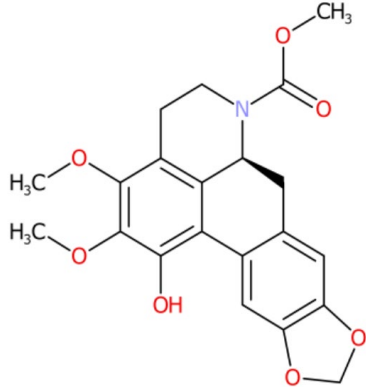
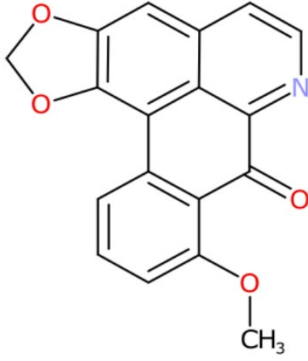
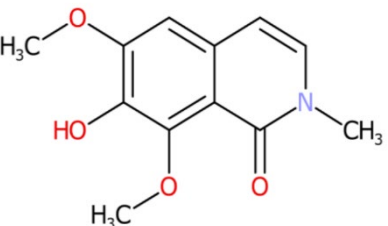
A partir do banco de 1.297 alcaloides, após o tratamento dos dados e aplicação dos critérios de exclusão, 980 estruturas foram submetidas à triagem virtual utilizando o modelo QSAR desenvolvido para a classe AB. Inicialmente, aplicou-se ao domínio de aplicabilidade (DA) para garantir que as predições fossem realizadas apenas para compostos dentro do espaço químico coberto pelo modelo, evitando extrapolações não confiáveis (Horvath; Marcou; Varnek, 2010; Sahigara *et al.*, 2012). Das 980 estruturas analisadas, 922 (94,1%) foram consideradas dentro do domínio de aplicabilidade, atestando a adequação do banco de dados ao espaço químico do modelo.

Utilizando como critério de seleção compostos classificados como ativos (classe A) com confiabilidade mínima de 85%, foram selecionadas 70 estruturas. A Tabela 7 apresenta as 10 primeiras estruturas ordenadas por maior percentual de confiabilidade, juntamente com seus números CAS, classes químicas e estruturas moleculares.

Tabela 7: 10 estruturas selecionadas a partir de maior confiabilidade de atividade na Triagem Virtual.

Identificador CAS	Nome Comum	Estrutura química	Classificação	Porcentual de confiabilidade
CAS900	Desacetilaspidoespermina		Indólico	96.00%
CAS1041	O-Metilisopilina		N-Óxido De Aporfina	96.00%
CAS623	R-Lotusina		Isoquinolínico	93.00%

CAS634	1,2-Dehydroreticulina		Isoquinolínico	93.00%
CAS647	Palmatina		Isoquinolínico	93.00%
CAS665	Oxogalucina		Oxoaporfinico	93.00%
CAS701	1,2,9-trimetoxi-10-hydroxiaporfina		Aporfinico	93.00%

CAS807	(S)-N-metoxicarbonil-3-metoxinordomesticina		Aporfinico	93.00%
CAS964	Oxoestefanina		Oxoaporfinico	93.00%
CAS967	Cherianoine		Isoquinolínico	93.00%

Fonte: Autor.

5.3.1 Análise da triagem virtual

A seguir, será apresentada uma análise detalhada dos dez alcaloides selecionados com maior percentual de confiabilidade predita para o modelo QSAR, relacionando-os com dados disponíveis na literatura acerca de suas fontes naturais, atividades biológicas conhecidas e, se existentes evidências de atividade contra o *F. oxysporum* ou fungos correlatos.

Um resultado particularmente relevante da triagem virtual foi a seleção da palmatina (CAS647), um alcaloide isoquinolínico com 93% de confiabilidade de atividade contra *F. oxysporum*. A palmatina é um composto natural isolado de diversas espécies vegetais,

incluindo *Tinospora capillipes* (Deng; Zhang; Luo., 2012, tubérculos de *Stephania dielsiana* (Thien *et al.*, 2018), e sua atividade antifúngica já foi demonstrada experimentalmente na literatura. Deng; Zhang; Luo, (2012) isolaram palmatina de *Tinospora capillipes* e avaliaram sua atividade contra diversos fitopatógenos, incluindo *Fusarium oxysporum f. sp. niveum*. Os autores reportaram valores de EC50 entre 0,0348 e 0,8356 g/L, confirmando o potencial antifúngico deste alcaloide. Este achado é particularmente significativo pois ele valida experimentalmente a predição do modelo QSAR, demonstrando que a abordagem *in silico* foi capaz de identificar corretamente um composto com atividade comprovada.

Para os outros alcaloides selecionados, não foram encontrados testes para o *F. oxysporum*. Entretanto, informações importantes sobre os compostos selecionados com evidências indiretas ou contextual serão mencionadas a seguir.

- a) Cherianoine (CAS967) e o extrato de *Annona cherimola*: Embora a cherianoine isoladamente não tenha sido testada contra *F. oxysporum*, Trindade *et al.* (2025) demonstraram que extratos de óleo de resíduos biológicos de sementes e cascas de *Annona cherimola* Mill. apresentaram 60% de ação antifúngica no terceiro dia contra o patógeno. A cherianoine, isolada dos caules da mesma espécie, pode ser um dos componentes responsáveis por esta atividade, hipótese que merece investigação futura.
- b) Lotusina (CAS623) e o papel dos monobenzilisoquinolínicos na defesa vegetal: Lee, Hagel e Facchini (2013) mencionam que alcaloides monobenzilisoquinolínicos, classe à qual pertence a lotusina, são os principais componentes do lótus (*Nelumbo nucifera*) e, por possuírem alto valor nutricional especializado, podem desempenhar papel fundamental na defesa contra herbívoros e patógenos de plantas. Esta observação menciona suporte indireto ao potencial antifúngico da lotusina, sugerindo que sua presença em plantas pode estar associada a mecanismos de resistência a fitopatógenos, incluindo fungos do gênero *Fusarium*.
- c) Alcaloides de Aspidosperma e a atividade antimicrobiana: A desacetilaspidospermina (CAS900) é um representante típico dos alcaloides indólicos do gênero *Aspidosperma* (Pereira *et al.*, 2007), amplamente estudado por suas atividades biológicas e presente nas plantas *Aspidosperma neblinae* (Thomas; Schnoes; Biemann, 1969); *Aspidosperma polyneuron* (Bolzani *et al.*, 1987); *Aspidosperma pyriformium* (Mitaine *et al.*, 1996) e *Aspidosperma quebracho-blanco* (Lyon *et al.*, 1973). Embora não haja relatos específicos para *F. oxysporum*, a literatura sobre alcaloides indólicos em geral aponta para potenciais antimicrobianos e antifúngicos, justificando a investigação experimental deste composto.

Um resultado importante a ser mencionado é o caso da (S)-N-metoxicarbonil-3-metoxinordomesticina (CAS807). Este alcaloide aporfínico, isolado de *Ocotea macrophylla*, foi testado por Pabon e Cuca (2010) contra *Fusarium oxysporum* f. sp. *lycopersici* e não apresentou atividade antifúngica na concentração avaliada (250 µg/µL). Os autores sugerem que a presença de substituintes eletronegativos no átomo de nitrogênio pode diminuir a atividade antifúngica.

Para os demais alcaloides de CAS1041, CAS634, CAS665, CAS701 e CAS964, não foram encontrados relatos de atividade contra *F. oxysporum* ou mesmo contra outros fungos do gênero *Fusarium*. No entanto, suas classes químicas (aporfínicos, oxoaporfínicos) são reconhecidas por atividades biológicas diversas, incluindo efeitos anticancerígenos, anti-inflamatórios e, em alguns casos, antimicrobianos (Makarasen *et al.*, 2011; Azamov *et al.*, 2023). A oxoestefanina (CAS964), por exemplo, apresenta potente atividade contra células tumorais e contra *Plasmodium falciparum* (Makarasen *et al.*, 2011), sugerindo que sua reatividade biológica pode se estender a outros organismos, incluindo fungos. A oxogalucina (CAS665) também tem atividades anti-inflamatórias e anticancerígenas reportadas (Azamov *et al.*, 2023). Estas observações, embora indiretas, reforçam a relevância de investigar experimentalmente o potencial antifúngico destes compostos.

6. CONCLUSÃO

A presente dissertação alcançou seu objetivo principal ao desenvolver e validar um modelo QSAR preditivo, robusto e generalizável para a identificação de alcaloides com potencial atividade contra *Fusarium oxysporum*. O modelo demonstrou excelente desempenho em todas as etapas de validação (acurácia > 86%, $\kappa > 0,74$, AUC > 0,92) e foi capaz de selecionar, com alta confiabilidade, 70 alcaloides candidatos, incluindo a palmatina – cuja atividade antifúngica contra *F. oxysporum* é experimentalmente confirmada na literatura, validando externamente a abordagem empregada.

Além disso, a análise dos descritores moleculares forneceu informações sobre as características estruturais necessárias para a atividade antifúngica, contribuindo para o avanço do conhecimento na área e para o planejamento racional de novos fitofarmacêuticos.

Os resultados obtidos posicionam esta pesquisa como uma contribuição relevante para a quimioinformática aplicada à fitossanidade, oferecendo uma ferramenta preditiva confiável que pode acelerar e otimizar o processo de descoberta de novos antifúngicos de origem natural, contribuindo para o manejo sustentável da fusariose e para a redução da dependência de fungicidas sintéticos convencionais.

REFERÊNCIAS BIBLIOGRÁFICAS

AFANTITIS, Antreas *et al.* Ligand - Based virtual screening procedure for the prediction and the identification of novel β -amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. **European Journal of Medicinal Chemistry**, v. 46, n. 2, p. 497–508, fev. 2011.

Anaconda Software Distribution. Anaconda Inc., nov. 2016. Disponível em: <https://anaconda.com/>

ANDRADE, Jônathas G.; FREITAS, Humberto F.; CASTILHO, Marcelo S. Estudos de QSAR 2D baseados em descritores topológicos e fragmentos moleculares para uma série de derivados azólicos ativos contra *Candida albicans*. **Química Nova**, v. 35, n. 3, p. 466–472, 2012.

ANYUBAGA, Salim Bitrus *et al.* Chemo-informatics applications in the design of novel 7-keto-sempervirolo derivatives as SmCB1 inhibitors with potential for treatment of Schistosomiasis. **Heliyon**, v. 10, n. 1, p. e23115, jan. 2024.

AZAMOV, Bakhovuddin *et al.* Oxoglaucone Suppresses Hepatic Fibrosis by Inhibiting TGF β -Induced Smad2 Phosphorylation and ROS Generation. **Molecules**, v. 28, n. 13, 1 jul. 2023.

BAI, G.; SU, Z.; CAI, J. Wheat resistance to *Fusarium* head blight. **Canadian Journal of Plant Pathology**, v. 40, n. 3, p. 336–346, 3 jul. 2018.

BAJŽELJ, Benjamin; NOVIČ, Marjana; DRGAN, Viktor. Utilizing Molecular Descriptor Importance to Enhance Endpoint Predictions. **Toxics**, v. 13, n. 5, p. 383, 9 maio 2025.

BAO, L.-Q. *et al.* Development of Activity Rules and Chemical Fragment Design for In Silico Discovery of AChE and BACE1 Dual Inhibitors against Alzheimer's Disease. **Molecules**, v. 28, n. 8, p. 3588, 20 abr. 2023.

BASAK, Subhash C. *et al.* Quantitative structure-activity relationship modeling of juvenile hormone mimetic compounds for *Culex pipiens* larvae, with a discussion of descriptor-thinning methods. *In: American Chemical Society*, 2006.

BASKIN, I. I.; KIREEVA, N.; VARNEK, A. The one-class classification approach to data description and to models applicability domain. **Molecular Informatics**, v. 29, n. 8–9, p. 581–587, 17 set. 2010.

BAUMANN, Désirée; BAUMANN, Knut. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. **Journal of Cheminformatics**, v. 6, n. 1, 2014.

BERNAL, F. A.; COY-BARRERA, E. Composition and Antifungal Activity of the Alkaloidal Fraction of *Lupinus mirabilis* Leaves: A Biochemometrics-Based Exploration. **Molecules**, v. 27, n. 9, p. 2832, 29 abr. 2022.

BERTHOLD, M. R. *et al.* KNIME: The Konstanz Information Miner. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, 2008. p. 319–326.

BHAMBHANI, S.; KONDHARE, K. R.; GIRI, A. P. **Diversity in chemical structures and biological properties of plant alkaloids**. *Molecules* MDPI AG, 1 jun. 2021.

BOLZANI, Vanderlan Da S. *et al.* Indole alkaloid evolution in Aspidosperma. *Biochemical Systematics and Ecology*, v. 15, n. 2, p. 187–200, mar. 1987.

BORREGO-MUÑOZ, Paola; OSPINA, Felipe; QUIROGA, Diego. **A compendium of the most promising synthesized organic compounds against several fusarium oxysporum species: Synthesis, antifungal activity, and perspectives**. *Molecules* MDPI AG, , 1 jul. 2021.

BOSC, N. *et al.* Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *Journal of Cheminformatics*, v. 11, n. 1, 10 jan. 2019.

BREIMAN, Leo *et al.* **Classification And Regression Trees**. Routledge, 2017.

BREIMAN, Leo. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, out. 2001.

CARHART, R. E.; SMITH, D. H.; VENKATARAGHAVAN, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *Journal of Chemical Information and Computer Sciences*, v. 25, n. 2, p. 64–73, 1 maio 1985.

CARPENTER, K. A. *et al.* Deep Learning and Virtual Drug Screening. *Future Medicinal Chemistry*, v. 10, n. 21, p. 2557–2567, 5 nov. 2018.

CARPENTER, K. A.; HUANG, X. Machine Learning-based Virtual Screening and Its Applications to Alzheimer’s Drug Discovery: A Review. *Current Pharmaceutical Design*, v. 24, n. 28, p. 3347–3358, 7 dez. 2018.

CARRASCO CABRERA, Luis *et al.* The 2021 European Union report on pesticide residues in food. *EFSA Journal*, v. 21, n. 4, 1 abr. 2023.

CELY-VELOZA, Willy *et al.* Antifungal activity against *Fusarium oxysporum* of quinolizidines isolated from three controlled-growth Genisteeae plants: structure–activity relationship implications. *Natural Products and Bioprospecting*, v. 13, n. 1, p. 9, 20 dez. 2023.

CHAWLA, N. v. *et al.* SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 1 jun. 2002.

CHEBOTAEV, Platon P. *et al.* Predicting fluorescence to singlet oxygen generation quantum yield ratio for BODIPY dyes using QSPR and machine learning. *Physical Chemistry Chemical Physics*, v. 26, n. 38, p. 25131–25142, 2024.

- CHEN, B. *et al.* Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: A review of methods and resources. **Journal of Hazardous Materials**. Elsevier, 5 dez. 2015.
- CHEN, Chung-Yi *et al.* Four alkaloids from *Annona cherimola*. **Phytochemistry**, v. 56, n. 7, p. 753–757, abr. 2001.
- CHEN, R.-C. *et al.* Selecting critical features for data classification based on machine learning methods. **Journal of Big Data**, v. 7, n. 1, p. 52, 23 dez. 2020.
- COHEN, Jacob. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1 abr. 1960.
- COLEMAN, J. J. The *Fusarium solani* species complex: Ubiquitous pathogens of agricultural importance. **Molecular Plant Pathology**, v. 17, n. 2, p. 146–158, 1 fev. 2016.
- DAI, R. *et al.* Anti-Alzheimer's disease potential of traditional chinese medicinal herbs as inhibitors of BACE1 and AChE enzymes. **Biomedicine & Pharmacotherapy**, v. 154, p. 113576, out. 2022.
- DAMALE, M. *et al.* Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. **Mini-Reviews in Medicinal Chemistry**, v. 14, n. 1, p. 35–55, 17 jan. 2014.
- DAVIS, Jesse; GOADRICH, Mark. **The relationship between precision-recall and ROC curves**. 2006.
- DE CHAVES, Magda Antunes *et al.* **Fungicide Resistance in *Fusarium graminearum* Species Complex**. **Current Microbiology** Springer, , 1 fev. 2022.
- DENG, Yecheng; ZHANG, Ming; LUO, Haiyu. Identification and antimicrobial activity of two alkaloids from traditional Chinese medicinal plant *Tinospora capillipes*. **Industrial Crops and Products**, v. 37, n. 1, p. 298–302, maio 2012.
- DUCHOWICZ, Pablo R. *et al.* QSAR analysis for heterocyclic antifungals. **Bioorganic & Medicinal Chemistry**, v. 15, n. 7, p. 2680–2689, abr. 2007.
- EDEL-HERMANN, V.; LECOMTE, C. Current Status of *Fusarium oxysporum* *Formae Speciales* and Races. **Phytopathology®**, v. 109, n. 4, p. 512–530, abr. 2019.
- ERIKSSON, Jakob. 42. Jakob Eriksson: Ueber die Specialisirung des Parasitismus bei den Getreiderostpilzen. **Berichte der Deutschen Botanischen Gesellschaft**, v. 12, n. 9, p. 292–331, 21 dez. 1894.
- EUROPEAN FOOD SAFETY AUTHORITY. The 2010 European Union Report on Pesticide Residues in Food. **EFSA Journal**, v. 11, n. 3, 1 mar. 2013.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, jun. 2006.

FERNANDES, A. DE O. *et al.* Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening. **Drugs and Drug Candidates**, v. 4, n. 3, p. 33, 4 jul. 2025.

FERREIRA, Leticia Tiburcio *et al.* QSAR-Based Virtual Screening of Natural Products Database for Identification of Potent Antimalarial Hits. 2021.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of Chemical Information and Modeling**. American Chemical Society, 26 jul. 2010.

FRISCHE, T. *et al.* **5-Point programme for sustainable plant protection**. **Environmental Sciences Europe**. Springer Verlag, 1 dez. 2018.

GASPAROTTO, L. *et al.* **Fusarium oxysporum f. sp. cubense raça 4 tropical: perigo para a bananicultura nacional**. 2020. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1128453>>.

GASTEIGER, J.; ENGEL, T. **Chemoinformatics**. Wiley eBooks, 25 set. 2003.

GHINI, R.; KIMATI, H. **Resistência de fungos a fungicidas**. São Paulo: Embrapa Meio Ambiente, 2000.

GOERIG, M.; SCHULTE AM ESCH, J. [Friedrich Wilhelm Adam Sertürner—the discoverer of morphine]. **Anesthesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie: AINS**, v. 26, n. 8, p. 492–498, 1 dez. 1991.

GOLBRAIKH, A. *et al.* Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment. In: **Handbook of Computational Chemistry**. Dordrecht: Springer Netherlands, 2016. p. 1–48.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 2016. Disponível em: <<https://www.deeplearningbook.org>>.

GORDON, T. R. *Fusarium oxysporum* and the *Fusarium* Wilt Syndrome. **Annual Review of Phytopathology**, p. 15, 2025.

GOYAL, S. Ecological role of alkaloids. In: **Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes**. Springer Berlin Heidelberg, 2013. p. 149–171.

GRAMATICA, P. Principles of QSAR Modeling. **International Journal of Quantitative Structure-Property Relationships**, v. 5, n. 3, p. 61–97, 1 jul. 2020.

GREEN, D.M. and Swets, J.A. **Signal Detection Theory and Psychophysics**. Wiley, New York. 1966.

GRISONI, Francesca *et al.* Matrix-based Molecular Descriptors for Prospective Virtual Compound Screening. **Molecular Informatics**, v. 36, n. 1, 1 jan. 2017.

GRISONI, Francesca *et al.* Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach. *In: Methods in Molecular Biology*. [S.l.]: Humana Press, New York, NY, 2018. v. 1800 p. 3–53.

GUTIÉRREZ-GRIJALVA, Erick Paul *et al.* Plant alkaloids: Structures and bioactive properties. *In: Plant-derived Bioactives: Chemistry and Mode of Action*. [S.l.]: Springer Singapore, 2020. p. 85–117.

HALL, L. H.; KIER, L. B. Electrotological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. **Journal of Chemical Information and Computer Sciences**, v. 35, n. 6, p. 1039–1045, 1 nov. 1995.

HANSCH, CORWIN.; FUJITA, TOSHIO. p - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. **Journal of the American Chemical Society**, v. 86, n. 8, p. 1616–1626, 1 abr. 1964.

HAO, Chenyang *et al.* Visualization and identification of benzyloisoquinoline alkaloids in various *Nelumbo nucifera* tissues. **Heliyon**, v. 9, n. 6, 1 jun. 2023.

HEMMATEENEJAD, B. *et al.* QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. **Chemometrics and Intelligent Laboratory Systems**, v. 64, n. 1, p. 91–99, out. 2002.

HIRATA, Kazumasa *et al.* 1,2-Dehydroreticuline synthase, the branch point enzyme opening the morphinan biosynthetic pathway. **Phytochemistry**, v. 65, n. 8, p. 1039–1046, abr. 2004.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. The MIT Press, 1992.

HORVATH, D.; MARCOU, G.; VARNEK, A. A unified approach to the applicability domain problem of QSAR models. **Journal of Cheminformatics**, v. 2, n. S1, maio 2010.

IBGE. **Produção Agrícola Municipal 2023**. Rio de Janeiro. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e->>.

IBRAHIM, Muhammad Tukur *et al.* Quantitative structure-activity relationship, molecular docking, drug-likeness, and pharmacokinetic studies of some non-small cell lung cancer therapeutic agents. **Beni-Suef University Journal of Basic and Applied Sciences**, v. 9, n. 1, p. 49, 10 dez. 2020.

IMAN, M.; DAVOOD, A.; BANAROU EI, N. QSAR study of chalcone derivatives as anti-Leishmania agents. **Turkish Journal of Chemistry**, v. 38, n. 5, p. 716–724, 17 set. 2014.

JOHNS, L. E. *et al.* Emerging health threat and cost of *Fusarium* mycotoxins in European wheat. **Nature Food**, v. 3, n. 12, p. 1014–1019, 1 dez. 2022.

KOTU, Vijay; DESHPANDE, Bala. Model Evaluation. *In: Data Science*. Elsevier, 2019. p. 263–279.

KRYGOWSKI, Tadeusz Marek; CYRAŃSKI, Michał Ksawery. **Structural aspects of aromaticity**. **Chemical Reviews**, maio 2001.

LANDIS, J. Richard; KOCH, Gary G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159, mar. 1977.

LEARDI, R.; BOGGIA, R.; TERRILE, M. Genetic algorithms as a strategy for feature selection. **Journal of Chemometrics**, v. 6, n. 5, p. 267–281, 1992.

LEE, Eun-Jeong; HAGEL, Jillian M.; FACCHINI, Peter J. Role of the phloem in the biochemistry and ecophysiology of benzyloisoquinoline alkaloid metabolism. **Frontiers in Plant Science**, v. 4, n. JUN, 11 jun. 2013.

LI, Ming; GAO, Qian; YU, Tianfei. **Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters**. **BMC CancerBioMed Central Ltd**, , 1 dez. 2023.

LI, Yongchun; ZHAO, Meirong; ZHANG, Zhi. Quantitative proteomics reveals the antifungal effect of canthin-6-one isolated from *Ailanthus altissima* against *Fusarium oxysporum* f. sp. *cucumerinum* in vitro. **PLoS ONE**, v. 16, n. 4 April, 1 abr. 2021.

LIEVENS, B. *et al.* Evolutionary relationships between *Fusarium oxysporum* f. sp. *lycopersici* and *F. oxysporum* f. sp. *radicis-lycopersici* isolates inferred from mating type, elongation factor-1 α and exopolygalacturonase sequences. **Mycological Research**, v. 113, n. 10, p. 1181–1191, out. 2009.

LIMA, Felipe Tomazelli; SOUZA, Vinicius M. A. A Large Comparison of Normalization Methods on Time Series. **Big Data Research**, v. 34, p. 100407, nov. 2023.

LOMBARD, L. *et al.* Epitypification of *Fusarium oxysporum* – Clearing the taxonomic chaos. **Persoonia: Molecular Phylogeny and Evolution of Fungi**, v. 43, p. 1–47, 1 dez. 2019.

LOPES, C. A.; REIS, A. **Diferenciando asurchas do tomateiro**. Brasília. 2022. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1151639>>.

LUSTED, Lee B. Signal Detectability and Medical Decision-Making. **Science**, v. 171, n. 3977, p. 1217–1219, 26 mar. 1971.

LYON, R. L. *et al.* Biological and Phytochemical Evaluation of Plants XI: Isolation of Aspidospermine, Quebrachidine, Rhazinilam, (—)-Pyrifolidine, and Akuammidine from *Aspidosperma Quebracho-blanco* (Apocynaceae). **Journal of Pharmaceutical Sciences**, v. 62, n. 2, p. 218–221, fev. 1973.

MAKARASEN, Arthit *et al.* Cytotoxic and antimicrobial activities of aporphine alkaloids isolated from *stephania venosa* (Blume) spreng. **Planta Medica**, v. 77, n. 13, p. 1519–1524, 2011.

MANICI, L. M.; CAPUTO, F.; SACCA, M. L. Secondary metabolites released into the rhizosphere by *Fusarium oxysporum* and *Fusarium spp.* as underestimated component of nonspecific replant disease. **Plant and Soil**, v. 415, n. 1–2, p. 85–98, 1 jun. 2017.

MAURI, A. alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. **Methods in pharmacology and toxicology**, p. 801–820, 1 jan. 2020.

MAURI, A.; CONSONNI, V.; TODESCHINI, R. Molecular Descriptors. In: **Handbook of Computational Chemistry**. Cham: Springer International Publishing, 2017. p. 2065–2093.

MAURI, Andrea; BERTOLA, Matteo. Alvascience: A New Software Suite for the QSAR Workflow Applied to the Blood–Brain Barrier Permeability. **International Journal of Molecular Sciences**, v. 23, n. 21, 1 nov. 2022.

MCHUGH, Mary L. Interrater reliability: the kappa statistic. **Biochemia medica**, v. 22, n. 3, p. 276–82, 2012.

MELAGRAKI, G. *et al.* In silico exploration for identifying structure-activity relationship of MEK inhibition and oral bioavailability for isothiazole derivatives. **Chemical Biology and Drug Design**, v. 76, n. 5, p. 397–406, nov. 2010.

MICHEREFF, Sami J.; ANDRADE, Domingos E. G. T.; MENEZES, Maria. **Ecologia e Manejo de Patógenos Radiculares em Solos Tropicais**. Recife: UFRPE, 2005.

Ministério da agricultura, pecuária e. Abastecimento superintendência federal de agricultura, pecuária e. Abastecimento em são paulo –. Sfa-sp/mapa. **Ministério da agricultura, pecuária e abastecimento superintendência federal de agricultura, pecuária e diálogos para prevenção da raça 4 tropical da fusariose em bananeiras missão do mapa: promover o desenvolvimento sustentável das cadeias produtivas agropecuárias, em benefício**. [s.l.: s.n.]. Disponível em: <[https:// www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/cesesp/publicacoes/livros/ dialogos-prevencao-da-fusariose-em-bananeiras](https://www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/cesesp/publicacoes/livros/dialogos-prevencao-da-fusariose-em-bananeiras)>.

MITAINE, A. C. *et al.* Alkaloids from *Aspidosperma* Species from Bolivia. **Planta Medica**, v. 62, n. 05, p. 458–461, 4 out. 1996.

MORE, S. J. *et al.* Guidance Document on Scientific criteria for grouping chemicals into assessment groups for human risk assessment of combined exposure to multiple chemicals. **EFSA Journal**, v. 19, n. 12, 1 dez. 2021.

MOUSSA, J. E.; STEWART, J. J. P. **MOPAC software ver. 22.1.1.**, 29 jan. 2024.

MUKHERJEE, R. K.; KUMAR, V.; ROY, K. Chemometric modeling of plant protection products (PPPs) for the prediction of acute contact toxicity against honey bees (*A. mellifera*): A 2D-QSAR approach. **Journal of Hazardous Materials**, v. 423, p. 127230, fev. 2022.

MUNKVOLD, Gary P. Fusarium species and their associated mycotoxins. *In: Methods in Molecular Biology*. [S.l.]: Humana Press Inc., 2017. v. 1542 p. 51–106.

NAHM, Francis Sahngun. Receiver operating characteristic curve: overview and practical use for clinicians. **Korean Journal of Anesthesiology**, v. 75, n. 1, p. 25–36, 1 fev. 2022.

NAQVI, Syed Atif Hasan *et al.* **Fungicide resistance in Fusarium species: exploring environmental impacts and sustainable management strategies**. **Archives of Microbiology** Springer Science and Business Media Deutschland GmbH, , 1 fev. 2025.

NAVARRO, V. R. *et al.* Alkaloids from *Duguetia flagellaris* Huber (Annonaceae). **Revista Brasileira de Plantas Mediciniais**, v. 3, n. 2, p. 23–29, 2 abr. 2001.

NEVES, B. J. *et al.* QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. **Frontiers in Pharmacology**, v. 9, 13 nov. 2018.

NEVES, Bruno J. *et al.* Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening. **Journal of Medicinal Chemistry**, v. 59, n. 15, p. 7075–7088, 11 ago. 2016.

O'BOYLE, N. M. *et al.* Open Babel: An Open chemical toolbox. **Journal of Cheminformatics**, v. 3, n. 10, out. 2011.

OECD. **Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models**. Paris: OECD, 2014.

PABON, Ludy Cristina; CUCA, Luis Enrique. Aporphine alkaloids from *Ocotea macrophylla* (Lauraceae). **Química Nova**, v. 33, n. 4, p. 875–879, 2010.

PAIXÃO, V. V. M. *et al.* Novel Antischistosomal Drug Targets: Identification of Alkaloid Inhibitors of SmTGR via Integrated In Silico Methods. **Pathogens**, v. 14, n. 6, 1 jun. 2025.

PANOV, P.; DŽEROSKI, S. Combining Bagging and Random Subspaces to Create Better Ensembles. *In: Advances in Intelligent Data Analysis VII*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 118–129.

- PARVANDEH, S. *et al.* Consensus features nested cross-validation. **Bioinformatics**, v. 36, n. 10, p. 3093–3098, 1 maio 2020.
- PEREIRA, Maria de M. *et al.* Alcalóides indólicos isolados de espécies do gênero *Aspidosperma* (Apocynaceae). **Química Nova**, v. 30, n. 4, p. 970–983, ago. 2007.
- POURBASHEER, E. *et al.* QSAR study of ACK1 inhibitors by genetic algorithm–multiple linear regression (GA–MLR). **Journal of Saudi Chemical Society**, v. 18, n. 5, p. 681–688, nov. 2014.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, 2021. Disponível em: <<https://www.R-project.org/>>. Acesso em: 4 out. 2024
- RÁCZ, A.; BAJUSZ, D.; HÉBERGER, K. Intercorrelation Limits in Molecular Descriptor Preselection for QSAR/QSPR. **Molecular Informatics**, v. 38, n. 8–9, 1 ago. 2019.
- RAHNENFÜHRER, Jörg *et al.* Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. **BMC Medicine**, v. 21, n. 1, p. 182, 15 maio 2023.
- RAPPE, A. K. *et al.* UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. **Journal of the American Chemical Society**, v. 114, n. 25, p. 10024–10035, 1 dez. 1992.
- RDKit: Open-source cheminformatics**. Disponível em: <<https://www.rdkit.org>>
- REUTLINGER, M. *et al.* Chemically Advanced Template Search (CATS) for Scaffold-Hopping and Prospective Target Prediction for ‘Orphan’ Molecules. **Molecular Informatics**, v. 32, n. 2, p. 133–138, 7 fev. 2013.
- RINKER, T. W.; KURKIEWICZ, D. **pacman: Package Management for R**. New York, 2018. Disponível em: <<http://github.com/trinker/pacman>>
- RODRIGUES, N. E. *et al.* QSAR-Guided Proposition of N-(4-methanesulfonyl) Benzoyl-N’-(Pyrimidin-2-yl) Thioureas as Effective and Safer Herbicides. **Bulletin of Environmental Contamination and Toxicology**, v. 108, n. 6, p. 1019–1025, 1 jun. 2022.
- RODRIGUES, R. P. *et al.* Virtual screening strategies in drug design. **Revista Virtual de Química**, v. 4, n. 6, p. 739–776, 2012.
- ROMÁN-MONTES, Carla M. *et al.* **Molecular Identification and Antifungal Susceptibility of *Fusarium* spp. Clinical Isolates**. *Mycoses*, v. 68, n. 1, 12 jan. 2025.
- RUDRAPAL, M.; CHETIA, D. Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme. **Journal of Drug Delivery and Therapeutics**, v. 10, n. 4, p. 225–233, 15 jul. 2020.

SAHIGARA, F. *et al.* Comparison of different approaches to define the applicability domain of QSAR models. **Molecules**, v. 17, n. 5, p. 4791–4810, maio 2012.

SANDHU, Hardeep; KUMAR, Rajaram Naresh; GARG, Prabha. Machine learning-based modeling to predict inhibitors of acetylcholinesterase. **Molecular Diversity**, v. 26, n. 1, p. 331–340, 23 fev. 2022.

SEVERO, R. *et al.* *Fusarium* species causing root rot and wilt in tomato in Brazil. **Journal of Phytopathology**, v. 172, n. 1, 3 jan. 2024.

SLIWOSKI, Gregory; MENDENHALL, Jeffrey; MEILER, Jens. Autocorrelation descriptor improvements for QSAR: 2DA_Sign and 3DA_Sign. **Journal of Computer-Aided Molecular Design**, v. 30, n. 3, p. 209–217, 31 mar. 2016.

SNYDER, W. C.; HANSEN, H. N. THE SPECIES CONCEPT IN *FUSARIUM*. **American Journal of Botany**, v. 27, n. 2, p. 64–67, fev. 1940.

SONEGO, P.; KOCSOR, A.; PONGOR, S. ROC analysis: applications to the classification of biological sequences and 3D structures. **Briefings in Bioinformatics**, v. 9, n. 3, p. 198–209, 11 jan. 2008.

STEWART, James J. P. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. **Journal of Molecular Modeling**, v. 19, n. 1, p. 1–32, jan. 2013.

TANNER, Wilson P.; SWETS, John A. A decision-making theory of visual detection. **Psychological Review**, v. 61, n. 6, p. 401–409, 1954.

TEIXEIRA, L. *et al.* Soil attributes, plant nutrition, and *Fusarium* wilt of banana in São Paulo, Brazil. **Tropical Plant Pathology**, v. 46, n. 4, p. 443–454, 25 mar. 2021.

TETZ, Victor *et al.* Antifungal activity of a novel synthetic polymer M451 against phytopathogens. **Frontiers in Microbiology**, v. 14, 2023.

THIEN, Dao Duc *et al.* Cytotoxic Alkaloids from *Stephania dielsiana*. **Chemistry of Natural Compounds**, v. 54, n. 3, p. 613–616, 1 maio 2018.

THOMAS, D. W.; SCHNOES, H. K.; BIEMANN, K. The alkaloids of *Aspidosperma neblinae*. An application of a directly coupled gas chromatograph-mass spectrometer. **Experientia**, v. 25, n. 7, p. 678–679, jul. 1969.

TIBSHIRANI, Robert. Regression Shrinkage and Selection Via the Lasso. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 58, n. 1, p. 267–288, 1 jan. 1996.

TODESCHINI, Roberto; CONSONNI, Viviana. **Handbook of Molecular Descriptors**. 2008. Disponível em: <<https://www.wiley.com/en-us/Handbook+of+Molecular+Descriptors-p-9783527613113#description-section>>.

TRAN, Thu-Hien *et al.* Dual roles of oxostephanine as an Aurora kinase inhibitor and angiogenesis suppressor. **International Journal of Molecular Medicine**, v. 50, n. 5, p. 133, 13 set. 2022.

TRINDADE, Clarissa *et al.* Inhibitory activity of the peel and seed of *Annona cherimola* Mill. against *Fusarium* and *Aspergillus* pathogens. **AIMS Agriculture and Food**, v. 10, n. 1, p. 155–176, 2025.

TROPSHA, A. **Best practices for QSAR model development, validation, and exploitation**. **Molecular Informatics** Wiley-VCH Verlag, 12 jul. 2010.

TZATZARAKIS, M. N. *et al.* Comparison of in vitro activities of amphotericin, clotrimazole, econazole, miconazole, and nystatin against *Fusarium oxysporum*. **Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes**, v. 36, n. 3, p. 331–340, 2001.

VAN LOMMEL, Ruben *et al.* A Quantum Chemical Deep-Dive into the π - π Interactions of 3-Methylindole and Its Halogenated Derivatives—Towards an Improved Ligand Design and Tryptophan Stacking. **Pharmaceuticals**, v. 15, n. 8, p. 935, 28 jul. 2022.

VAZHEV, V. *et al.* QSAR modeling of antifusarial activity of organic compounds. **Agrarian Bulletin of the**, v. 208, n. 05, p. 55–62, 29 jun. 2021.

VELOZA, Willy Fernando Cely; COY-BARRERA, Ericsson David. Development of a 3D-QSAR model from interaction energy maps-derived descriptors for the prediction of antifungal quinolizidines. In: 7TH BRAZILIAN CONFERENCE ON NATURAL PRODUCT/ XXXIII RESEM PROCEEDINGS, 2019, Rio de Janeiro. Anais eletrônicos, Galoá, 2019. Disponível em: <https://proceedings.science/bcnp-series/bcnp-2019/trabalhos/development-of-a-3d-qsar-model-from-interaction-energy-maps-derived-descriptors?lang=en>.

WANG, Yuqing *et al.* Cepharanthine Inhibits *Fusarium solani* via Oxidative Stress and CFEM Domain-Containing Protein Targeting. **Microorganisms**, v. 13, n. 6, 1 jun. 2025.

WILDMAN, Scott A.; CRIPPEN, Gordon M. Three-dimensional molecular descriptors and a novel QSAR method. **Journal of Molecular Graphics and Modelling**, v. 21, n. 3, p. 161–170, dez. 2002.

WILLETT, P. Chemoinformatics: A history. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 1, p. 46–56, jan. 2011.

WU, Y. *et al.* Recent advances for alkaloids as botanical pesticides for use in organic agriculture. **International Journal of Pest Management**. Taylor and Francis Ltd., 2021.

YIN, Y. *et al.* **Fungicide Resistance: Progress in Understanding Mechanism, Monitoring, and Management. Phytopathology.** American Phytopathological Society, 1 maio 2023.

ZDRAZIL, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. **Nucleic Acids Research**, v. 52, n. D1, p. D1180–D1192, 5 jan. 2024.

ZHANG, Chongqing *et al.* Potato glycoside alkaloids exhibit antifungal activity by regulating the tricarboxylic acid cycle pathway of *Fusarium solani*. **Frontiers in Microbiology**, v. 15, 15 abr. 2024.

ZHANG, S. *et al.* A novel Automated Lazy Learning QSAR (ALL-QSAR) approach: Method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. **Journal of Chemical Information and Modeling**, v. 46, n. 5, p. 1984–1995, 2006.

ZHANG, Xi-Lin *et al.* Alkaloids isolated from the endophytic fungus *Fusarium concentricum*. **Natural Product Research**, v. 38, n. 9, p. 1517–1523, 2 maio 2024.

ZHAO, B.; HE, D.; WANG, L. **Advances in *Fusarium* drug resistance research. Journal of Global Antimicrobial Resistance.** Elsevier Ltd, 1 mar. 2021.

ZHU, Z.; ZHAO, S.; WANG, C. β -Carboline Alkaloids from *Peganum harmala* Inhibit *Fusarium oxysporum* from *Codonopsis radix* through Damaging the Cell Membrane and Inducing ROS Accumulation. **Pathogens**, v. 11, n. 11, p. 1341, 13 nov. 2022.