



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

**ELAINE SANTOS DA CONCEIÇÃO**

**ESTUDO *IN SILICO* NA BUSCA DE ALCALOIDES  
POTENCIALMENTE INIBIDORES DE *Staphylococcus aureus***

***IN SILICO STUDY IN THE SEARCH FOR ALKALOIDS  
POTENTIALLY INHIBITORS OF Staphylococcus aureus***





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

**ELAINE SANTOS DA CONCEIÇÃO**

**ESTUDO *IN SILICO* NA BUSCA DE ALCALOIDES  
POTENCIALMENTE INIBIDORES DE *Staphylococcus aureus***

Dissertação de Mestrado apresentada  
ao Programa de Pós-Graduação em  
Química, da Universidade Federal de  
Sergipe, para a obtenção do título de  
Mestre em Química.

**Orientador: Prof. Dr. Tiago Branquinho Oliveira**

***IN SILICO STUDY IN THE SEARCH FOR ALKALOIDS  
POTENTIALLY INHIBITORS OF Staphylococcus aureus***

*Master's dissertation presented to the  
Postgraduate Program in Chemistry,  
Federal University of Sergipe to obtain  
the title of Master in Chemistry.*



FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL  
UNIVERSIDADE FEDERAL DE SERGIPE

C744e Conceição, Elaine Santos da  
Estudo in silico na busca de alcaloides potencialmente  
inibidores de *Staphylococcus aureus* / Elaine Santos da  
Conceição; orientador Tiago Branquinho Oliveira - São Cristóvão,  
2023.  
117 f. : il.

Dissertação (mestrado em Química) – Universidade Federal de  
Sergipe, 2023.

1. Inibidores enzimáticos. 2. Alcaloides. 3. Mastite. 4.  
Modelos matemáticos. 5. Moléculas - Modelos. I. Oliveira, Tiago  
Branquinho orient. II. Título.

CDU 543.6



**SERVIÇO PÚBLICO FEDERAL**  
**MINISTÉRIO DA EDUCAÇÃO**  
**UNIVERSIDADE FEDERAL DE SERGIPE**  
Programa de Pós-Graduação em Química  
PPGQ



**FOLHA DE APROVAÇÃO**

Membros da Comissão Julgadora da Dissertação de Mestrado de Elaine Santos da Conceição apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Sergipe em 31/10/2023.



Documento assinado digitalmente  
**TIAGO BRANQUINHO OLIVEIRA**  
Data: 31/10/2023 12:27:47-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Tiago Branquinho Oliveira  
Departamento de Farmácia - UFS



Documento assinado digitalmente  
**JOSE DIOGO DE LISBOA DUTRA**  
Data: 31/10/2023 12:30:48-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. José Diogo de Lisboa Dutra  
Departamento de Química - UFS

---

Dr. Miller Santos Ferreira  
Universidade



Documento assinado digitalmente  
**MILLER SANTOS FERREIRA**  
Data: 31/10/2023 12:42:33-0300  
Verifique em <https://validar.iti.gov.br>

*Dedico este trabalho aos meus pais, Maria José da Conceição e Erisvaldo Santos da Conceição, por todo esforço e sacrifício que tiveram para me dar a oportunidade de seguir meus sonhos e realizar minhas conquistas. E aos meus irmãos por todo incentivo.*

## AGRADECIMENTOS

Agradeço, primeiramente, a Deus pelo dom da vida e por seu amor e cuidado incondicional;

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

Agradeço ao meu orientador Prof. Dr. Tiago Branquinho Oliveira, por toda compreensão e colaboração, por todos os momentos de orientação, pela disposição, paciência e prontidão em ajudar e esclarecer dúvidas;

Agradeço à Universidade Federal de Sergipe (UFS), especialmente, ao Programa de Pós-Graduação em Química (PPGQ) da UFS-Campus São Cristóvão pela oportunidade da realização do curso de mestrado;

Agradeço ao grupo de pesquisa do Laboratório de Alimentos e Bebidas (LAAB) do Departamento de Farmácia da UFS por toda colaboração durante o desenvolvimento do trabalho; a todos os colegas do grupo de pesquisa *Computer nided science* (CASE), agradeço em especial aos colegas Valéria Vieira Moura Paixão e Eduardo Borba Alves por toda colaboração e pelo conhecimento compartilhado. Aos colegas desta pós-graduação, Camila Goes, Felipe Gabriel e Luiz Felipe. Vocês foram de extrema importância para que tudo se concretizasse.

*“Tudo que um sonho precisa para ser realizado é alguém que  
acredite que ele possa ser realizado”.  
(Roberto Shinyashiki)*

## RESUMO

O *Staphylococcus aureus* é considerado uma espécie de bactéria invasiva e resistente a tratamentos e a causa de múltiplas infecções em humanos e animais. É o principal patógeno responsável pela mastite bovina, uma doença complexa e de difícil controle em rebanhos. Em geral, o tratamento das infecções causadas por bactérias apoia-se no uso de antibióticos. Porém, o *S. aureus* tem apresentado diversos mecanismos de resistência a muitos antibióticos, dificultando o tratamento das infecções. Assim, o desenvolvimento de novos medicamentos antiestafilocócicos, torna-se necessário. Nesse contexto, a diidrofolato redutase (DHFR), uma enzima essencial no crescimento e proliferação celular das bactérias, representa um alvo promissor no desenvolvimento de novos antibióticos. Com isso, o presente trabalho procura, a partir de técnicas *in silico*, selecionar alcaloides com atividade antiestafilocócica. Com este objetivo, dois bancos de dados foram selecionados da base de dados ChEMBL, um com estruturas testadas no *S. aureus* e outro com estruturas testadas na SaDHFR. As seleções de descritores para ambos os bancos foram exploradas com quatro algoritmos de aprendizado de máquina: *Linear Regression* (LR), *Pace Regression* (PR), *Support Vector Machine* (SVM) e *Multilayer Perceptron* (MLP). Os modelos construídos foram considerados aceitáveis, pois apresentaram valores de  $R^2$  e  $P^2 > 0,6$  e  $Q^2 > 0,5$ . Assim, os melhores modelos, com maior capacidade de previsão externa, selecionaram 30 alcaloides promissores em inibir a SaDHFR. Além disso, em uma segunda etapa, um estudo de *docking* molecular foi realizado no sítio ativo da SaDHFR para propor o modo de ligação destes alcaloides. Ramiflorinas A, um dos três alcalóides com maiores afinidades no sítio ativo da SaDHFR, já mostrou atividade inibitória *in vitro* contra *S. aureus*, conforme relatado na literatura. Com isso, conclui-se que os resultados encontrados neste trabalho podem contribuir para o desenvolvimento de novos candidatos, que deverão ser validados experimentalmente.

**Palavras-chave:** Triagem virtual. DHFR. Mastite bovina. QSAR. *Docking* molecular

## ABSTRACT

*Staphylococcus aureus* is considered a species of invasive and treatment-resistant bacteria and the cause of multiple infections in humans and animals. It is the main pathogen responsible for bovine mastitis, a complex disease that is difficult to control in herds. In general, the treatment of infections caused by bacteria relies on the use of antibiotics. Therefore, the development of new antistaphylococcal drugs becomes necessary. In this context, dihydrofolate reductase (DHFR), an essential enzyme in the growth and cell proliferation of bacteria, represents a promising target in the development of new antibiotics. Therefore, the present work seeks, using *in silico* techniques, to select alkaloids with antistaphylococcal activity. With this objective, two databases were selected from the ChEMBL database, one with structures tested in *S. aureus* and the other with structures tested in SaDHFR. Descriptor selections for both banks were explored with four machine learning algorithms: linear regression (LR), partial least squares regression (PLSR), support vector machine (SVM), and multilayer perceptron (MLP). The constructed models were considered acceptable, as they presented  $R^2$  and  $P^2$  values  $> 0.6$  and  $Q^2 > 0.5$ . Thus, the best models, with greater external prediction capacity, selected 30 promising alkaloids for inhibiting SaDHFR. Furthermore, in the second stage, a molecular docking study was carried out at the active site of SaDHFR to propose the binding mode of these alkaloids. Ramiflorine A, one of the three alkaloids with the highest affinities in the active site of SaDHFR, has already demonstrated *in vitro* inhibitory activity against *S. aureus*, as reported in the literature. Therefore, it is concluded that the results found in this work can contribute to the development of new candidates, which must be experimentally validated.

**Keywords:** Virtual screening. DHFR. Bovine mastitis. QSAR. Molecular Docking

## LISTA DE FIGURAS

|                                                                                                                                        |    |
|----------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1:</b> Estrutura do folato e suas três subestruturas. (A) pterina; (B) ácido para-aminobenzóico; (C) ácido glutâmico. ....   | 8  |
| <b>Figura 2:</b> Estrutura da DHFR do <i>S. aureus</i> na representação de fita. ....                                                  | 9  |
| <b>Figura 3:</b> Estruturas químicas dos inibidores de DHFR: iclaprim (A), trimetoprima (B). ....                                      | 10 |
| <b>Figura 4:</b> Esquema simplificado do processo de desenvolvimento do modelo QSAR. ....                                              | 15 |
| <b>Figura 5:</b> Esquema geral das etapas empregadas na identificação de alcaloides inididores de <i>S. aureus</i> . ....              | 20 |
| <b>Figura 6:</b> Etapas básicas realizadas pelo AG ....                                                                                | 24 |
| <b>Figura 7:</b> Importância relativa dos descritores selecionados pelo GImboost. ....                                                 | 29 |
| <b>Figura 8:</b> Matriz de correlação entre os descritores selecionados pelo AG para o SaDB. ....                                      | 31 |
| <b>Figura 9:</b> Matriz de correlação entre os descritores selecionados pelo AG para o SdhfrDB ((a) SdhfrDB-2D e (b) SdhfrDB-3D). .... | 31 |
| <b>Figura 10:</b> Relação dos valores reais vs os valores previstos para os modelos do SaDB. ....                                      | 34 |
| <b>Figura 11:</b> Relação dos valores reais vs os valores previstos para os modelos do SdhfrDB-2D. ....                                | 35 |
| <b>Figura 12:</b> Relação dos valores reais vs os valores previstos para os modelos do SdhfrDB-3D. ....                                | 35 |
| <b>Figura 13:</b> Relação dos valores previstos vs os resíduos dos modelos para os modelos do conjunto SaDB ....                       | 36 |
| <b>Figura 14:</b> Relação dos valores previstos vs os resíduos dos modelos para a matriz SdhfrDB-2D ....                               | 37 |
| <b>Figura 15:</b> Relação dos valores previstos vs os resíduos dos modelos para a matriz SdhfrDB-3D ....                               | 37 |
| <b>Figura 16:</b> Contribuição média dos descritores selecionados para as previsões do modelo MLP. ....                                | 40 |
| <b>Figura 17:</b> Correlação entre o descritor GATS5s vs pIC <sub>50</sub> . ....                                                      | 41 |
| <b>Figura 18:</b> Estruturas com os maiores valores do descritor GATS5s. ....                                                          | 42 |
| <b>Figura 19:</b> Estruturas com os menores valor do descritor GATS5s. ....                                                            | 42 |

|                                                                                                                                                                                                                                                                                                                     |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 20:</b> Estruturas com valores do descritor $nRNR_2$ , $nN(CO)_2$ , $nArCOOH$ , $nOht$ e $nOHp$ .                                                                                                                                                                                                         | 43 |
| <b>Figura 21:</b> Estruturas com valores do descritor $F04[N-O]$ , $B06[O-O]$ e $B09[O-O]$ .                                                                                                                                                                                                                        | 44 |
| <b>Figura 22:</b> Estruturas com valores altos do descritor $minsCH_3$ .                                                                                                                                                                                                                                            | 45 |
| <b>Figura 23:</b> Estruturas que apresentam valores altos do descritor $qpmax$ .                                                                                                                                                                                                                                    | 46 |
| <b>Figura 24:</b> Estruturas com valores do descritor $s3\_phSize$ .                                                                                                                                                                                                                                                | 47 |
| <b>Figura 25:</b> Estruturas que apresentam valores baixos e alto do descritor $SpMax\_Bh(p)$ . (Estrutura A e B apresentam valores baixos do descritor e estrutura C valor alto).                                                                                                                                  | 49 |
| <b>Figura 26:</b> Relação do $pIC_{50}$ com os descritores $SpMin1\_Bh(m)$ em (a) e relação do descritor $SpMax\_X$ em (b).                                                                                                                                                                                         | 50 |
| <b>Figura 27:</b> Relação do $pIC_{50}$ com o descritor $CATS2D\_04\_DL$ os pontos pretos indicam centros lipofílicos e azuis os centros doadores de hidrogênio que estão a 4 ligações de distância.                                                                                                                | 52 |
| <b>Figura 28:</b> Relação do $pIC_{50}$ com os descritores $CATS3D\_05\_DL$ e $CATS3D\_03\_LL$ . Os pontos pretos indicam centros lipofílicos e azuis doadores de hidrogênio.                                                                                                                                       | 54 |
| <b>Figura 29:</b> Relação dos descritores $WHALES100\_Rem$ e $SPH2$ para a atividade.                                                                                                                                                                                                                               | 55 |
| <b>Figura 30:</b> Estrutura 3D da melhor pose de cada alcaloide (à esquerda) e estrutura 2D com representação das interações (à direita). <b>a1/ a2)</b> Ramiflorines A, <b>b1/b2)</b> Aspidolimidine, <b>c1/c2)</b> Fendlerine, <b>d1/d2)</b> Strictosidine, <b>e1/e2)</b> Limaspermina, <b>f1/f2)</b> Cephaeline. | 64 |
| <b>Figura 31:</b> Estrutura 2D com representação das interações visualizadas para os dois alcaloides que apresentou os maiores valores scores. a) QN7 e b) Apateline.                                                                                                                                               | 67 |
| <b>Figura 32:</b> Domínio de aplicabilidade dos compostos do conjunto teste baseados nas distâncias euclidianas.                                                                                                                                                                                                    | 94 |
| <b>Figura 33:</b> Domínio de aplicabilidade dos compostos do conjunto de triagem baseados nas distâncias euclidianas.                                                                                                                                                                                               | 95 |

## LISTA DE TABELAS

|                                                                                                                                     |    |
|-------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabela 1:</b> Resultado do fator de inflação para a seleção SaDB. ....                                                           | 32 |
| <b>Tabela 2:</b> Resultado do fator de inflação dos descritores para o SadhfrDB .....                                               | 32 |
| <b>Tabela 3:</b> Parâmetros estatísticos obtidos a partir dos modelos desenvolvido para o conjunto SaDB. ....                       | 33 |
| <b>Tabela 4:</b> Parâmetros estatísticos obtidos a partir dos modelos desenvolvido para o conjunto SadhfrDB.....                    | 33 |
| <b>Tabela 5:</b> Resultados estatísticos dos Critérios de Aceitabilidade para os modelos. ....                                      | 38 |
| <b>Tabela 6:</b> Resultados obtidos no teste de randomização. ....                                                                  | 39 |
| <b>Tabela 7:</b> Número de estruturas que fazem parte do APD .....                                                                  | 56 |
| <b>Tabela 8:</b> Lista dos 15 alcaloides com maiores potencias de atividade previstos pelo consenso dos modelos LR-2D e SVM-3D..... | 57 |
| <b>Tabela 9:</b> Previsão do modo de ligação para os 15 alcaloides mais ativos obtidos pelo GOLD.....                               | 62 |

## LISTA DE ABREVIATURAS E SIGLAS

2D – Bidimensional

3D – Tridimensional

*ANN – Artificial Neural Networks* (Redes Neurais Artificiais)

*CADD – Computer-Assisted Drug Design* (planejamento de fármaco auxiliado por computador)

DA – Domínio de Aplicabilidade

DHFR – Dihidrofolato redutase

*DT – Decision tree* (Árvore de Decisão)

*HTS – High-throughput Screening* (Triagem de alto rendimento)

*KNN – K-Nearest Neighbors* (k-ésimo vizinho mais próximo)

*MAE – Mean Absolute Error* (Erro Absoluto Médio)

*MB – Mastite bovina*

*MLP – Multilayer Perceptron* (Perceptron Multicamadas)

*MLR – Multiple Linear Regression* (Regressão Linear Múltipla)

*NADPH – Nicotinamida-adenina dinucleotídeo fosfato*

PR – Pace Regression

QSAR – Quantitative structure–activity relationship (Relação Quantitativa entre Estrutura e Atividade)

*RF – Random Forest* (Floresta Randômica)

*RMSE – Root Mean Squared Error* (Erro Quadrático Médio)

*S. aureus – Staphylococcus aureus*

SaDHFR – Diidrofolato redutase do *Staphylococcus aureus*

SVM – *Support Vector Machine* (Máquina de Vetores de Suporte)

THF – Tetrahidrofolato

TMP – Trimetoprima

VS – *Virtual Screening* (Triagem Virtual)

## SUMÁRIO

|            |                                                                                     |           |
|------------|-------------------------------------------------------------------------------------|-----------|
| <b>1</b>   | <b>INTRODUÇÃO .....</b>                                                             | <b>1</b>  |
| <b>1.1</b> | <b>Mastite bovina .....</b>                                                         | <b>2</b>  |
| 1.1.1      | A mastite na pecuária brasileira .....                                              | 2         |
| 1.1.2      | Mastite bovina: etiologia e prevalência .....                                       | 3         |
| <b>1.2</b> | <b>Staphylococcus aureus .....</b>                                                  | <b>5</b>  |
| 1.2.1      | Tratamento .....                                                                    | 6         |
| 1.2.2      | Resistência antimicrobiana.....                                                     | 7         |
| <b>1.3</b> | <b>DHFR alvo molecular promissor para desenvolver antibacteriano.</b>               | <b>8</b>  |
| <b>1.4</b> | <b>Quimioinformática no planejamento de fármacos.....</b>                           | <b>11</b> |
| 1.4.1      | QSAR.....                                                                           | 12        |
| 1.4.2      | Docking Molecular.....                                                              | 15        |
| <b>2</b>   | <b>JUSTIFICATIVA .....</b>                                                          | <b>17</b> |
| <b>3</b>   | <b>OBJETIVOS .....</b>                                                              | <b>18</b> |
| 3.1        | Objetivo Geral .....                                                                | 18        |
| 3.2        | Objetivos Específicos .....                                                         | 18        |
| <b>4</b>   | <b>MATERIAL E MÉTODOS .....</b>                                                     | <b>19</b> |
| <b>4.1</b> | <b>Computador e Softwares .....</b>                                                 | <b>19</b> |
| <b>4.2</b> | <b>Métodos.....</b>                                                                 | <b>19</b> |
| 4.2.1      | Modelagem quantitativa de estrutura-atividade para S. aureus e SaDHFR.....          | 21        |
| 4.2.2      | Triagem virtual.....                                                                | 27        |
| 4.2.3      | Docking Molecular.....                                                              | 28        |
| <b>5</b>   | <b>RESULTADOS E DISCUSSÃO.....</b>                                                  | <b>29</b> |
| <b>5.1</b> | <b>Modelagem QSAR para identificação de alcaloides inibidores de S. aureus.....</b> | <b>29</b> |
| 5.1.1      | Cálculo e seleção dos descritores.....                                              | 29        |

|       |                                                                                                                           |    |
|-------|---------------------------------------------------------------------------------------------------------------------------|----|
| 5.1.2 | Construção dos modelos .....                                                                                              | 33 |
| 5.1.3 | Análise dos descritores .....                                                                                             | 39 |
| 5.2   | Triagem virtual baseada em ligantes .....                                                                                 | 55 |
| 5.3   | Docking Molecular .....                                                                                                   | 61 |
| 6     | CONCLUSÕES .....                                                                                                          | 68 |
| 7     | PERSPECTIVAS DO TRABALHO .....                                                                                            | 69 |
| 8     | REFERÊNCIAS .....                                                                                                         | 70 |
| 9     | APÊNDICES .....                                                                                                           | 89 |
| 9.1   | Apêndice 1- Descritores selecionados para construção dos modelos.....                                                     | 89 |
| 9.2   | Apêndice 2- Domínio de Aplicabilidade .....                                                                               | 94 |
| 9.4   | Apêndice 3- Seleção dos alcaloides potencialmente ativos pelo consenso da triagem virtual nos modelos LR-2D e SVM-3D..... | 96 |
| 9.5   | Apêndice 4- Referências das tabelas 11 e 12 .....                                                                         | 98 |

## 1 INTRODUÇÃO

A mastite bovina (MB) é um dos mais graves problemas nas fazendas de gado leiteiro em todo o mundo, a qual afeta a saúde do animal e gera perdas econômicas à indústria de laticínios. A MB apresenta grande prevalência nos rebanhos brasileiros e é responsável principalmente por queda na produção de leite, impacto negativo na qualidade do leite, descarte de leite contaminado e, ainda, gastos no tratamento e descarte dos animais cronicamente afetados [1].

É uma doença complexa, causada por um amplo espectro de patógenos (fungos, leveduras, algas, bactérias e vírus), entre os quais destacam-se como mais prevalentes, espécies de bactérias gram-positivas e gram-negativas [2]. Em posição de destaque, o *S. aureus* é considerado o principal patógeno relacionado à mastite. Em função da alta resistência a antimicrobianos, relacionada a seu potencial de disseminação e evasão do sistema imunológico do hospedeiro, e da gravidade dos quadros de infecção associados ao mesmo, é considerado o patógeno mais preocupante [3].

Os principais alvos que os fármacos atuam contra cepas de *S. aureus* incluem: Sortase A, uma enzima de membrana, central para virulência das bactérias e facilmente acessível aos antibióticos devido a sua presença na superfície bacteriana [4], DNA-girase é importante para replicação, recombinação e reparo do DNA [5], desidroesqualeno sintase é importante na síntese do pigmento estafiloxantina que protege o *S. aureus* e aumenta sua sobrevivência dentro do hospedeiro [6] e a diidrofolato redutase (DHFR), que está envolvida na via biossintética do folato, um precursor de bases nitrogenadas e alguns aminoácidos importantes para a sobrevivência da bactéria [7,8].

Esse último é considerado atualmente um importante alvo para descoberta de fármacos contra vários patógenos [9]. A classe de medicamentos dos antifolatos atua como bactericida, ou seja, leva à morte da bactéria [10]. Porém, medicamentos antes considerados eficientes na inibição de DHFR bacteriana têm apresentado ineficácia devido a mutações indutoras de resistências nas enzimas DHFR [11]. Além disso, já se sabe que o *S. aureus*

apresenta cepas resistentes a quase todos os mecanismos de atuação dos medicamentos disponíveis no mercado [12].

Nesse sentido, dado o crescimento de cepas multirresistentes em todo o mundo tal como a relevância de cepas resistentes de *S. aureus*, o principal obstáculo no tratamento de doenças infecciosas em animais e humanos é a descoberta de substâncias que possam apresentar atividade contra patógenos multirresistentes, como *S. aureus*, um desafio cada vez mais urgente e atual [9]. Desta forma, as técnicas de triagem auxiliadas por computadores apresentam importância significativa para contribuir na identificação de substâncias promissoras e *design* de novos medicamentos [13].

Produtos naturais da classe dos alcaloides têm sido alvo de investigação por apresentarem uma série de atividades farmacológicas, como anti-inflamatória, anticancerígena e antibacteriana. Tais substâncias vêm desempenhando papel importante para direcionar fármacos modernos para o tratamento de doenças infecciosas [7,9,14]. Assim, o presente trabalho objetiva a partir do uso de técnicas de triagem virtual, auxiliada por métodos computacionais, identificar substâncias promissoras da classe de alcaloides, como alternativa para direcionar o planejamento de fármacos no tratamento de doenças, causadas por *S. aureus*.

## **1.1 Mastite bovina**

### **1.1.1 A mastite na pecuária brasileira**

A pecuária leiteira no Brasil tem sido um dos mais relevantes segmentos do agronegócio, desempenhando um papel importante no processo de desenvolvimento social do país [15]. O Brasil está entre os maiores produtores mundial de leite, com uma produção aproximada em 25 bilhões de litros de leite [16].

Apesar da posição ocupada no cenário mundial, a produtividade média de leite no Brasil ainda é considerada baixa. Algumas das principais causas da baixa produtividade, além de fatores econômicos, reprodutivos e nutricionais, são as doenças infecciosas, com destaque para a mastite [17]. Essa doença

gera redução na produção e qualidade do leite, às vezes, de forma irreversível, podendo aumentar o desperdício de leite e óbito prematuro dos animais afetados, o que leva a perdas econômicas e produtivas tanto ao produtor de leite quanto à indústria de laticínios em todo o mundo [18].

Alterações na qualidade e redução da produção de leite são os problemas econômicos mais importantes ocasionados pela mastite na indústria leiteira. As alterações na qualidade do leite se devem ao aumento no número de células somáticas (CCS) que diminuem o tempo de vida de prateleira dos produtos lácteos, representando maior desperdício e menor rendimento de derivados lácteos. Além destes, as perdas decorrentes da mastite se devem a gastos no tratamento de quartos mamários infectados e serviços veterinários [19].

As infecções bacterianas que levam à mastite impõem um impacto negativo considerável na produção de leite, com perdas anuais estimadas em US\$ 30 bilhões somente nos EUA, devido aos custos com tratamento, morte do animal e diminuição da produção de leite [20]. No Brasil, as perdas econômicas decorrentes da mastite são estimadas em 4,6 bilhões de litros de leite [21].

Além das implicações financeiras apresentadas, a MB apresenta impactos relevantes na saúde pública, devido ao consumo do leite contaminado com agentes infecciosos que podem transmitir diversas doenças aos seres humanos, ou devido ao consumo de leite com presença de resíduos de antibióticos que são frequentemente usados no tratamento de vacas com mastite [15,18].

#### 1.1.2 *Mastite bovina: etiologia e prevalência*

Mastite, ou inflamação da glândula mamária, é uma enfermidade com etiologia complexa e multivariada que acomete boa parte do gado leiteiro, sua nomenclatura originou-se do grego *masto*- glândula e *itis*- inflamação. As principais causas de mastite estão relacionadas a fatores de interações entre animal, patógeno e ambiente [22].

Em função da intensidade e tipos dos sinais clínicos apresentados pelos animais infectados, a mastite é classificada como clínica ou subclínica. A mastite clínica (MC) apresenta alterações visíveis macroscopicamente no leite e/ou

úbere, o que facilita seu diagnóstico. As principais alterações observadas no úbere, são mamas endurecidas, inchaço, dor ao toque, vermelhidão e aumento da temperatura no animal. No leite, são encontrados sinais como aparecimento de grumos, pus ou sangue [18].

Já a mastite subclínica é caracterizada por alterações não visíveis de inflamação no úbere e a aparência do leite é macroscopicamente normal. Porém, é possível encontrar sinais como redução na produção de leite, alteração da qualidade do leite do rebanho pelo aumento de leucócitos (células somáticas), íons cloro, sódio, pH e diminuição na concentração de caseína, gorduras e lactose do leite [23].

A infecção no úbere ocorre majoritariamente por invasão de microrganismos geralmente durante ou após a ordenha, quando o canal do teto está aberto, facilitando a penetração dos microrganismos ambientais no úbere. A invasão pode ocorrer através das mãos do ordenhador ou pelo uso compartilhado de materiais infectados, principalmente quando os animais estão em ambientes contaminados durante o período entre as ordenhas [3].

Aproximadamente 200 agentes infecciosos são conhecidos como causadores da mastite bovina, os principais microrganismos associados incluem bactérias (maior prevalência), vírus e fungos [23]. Em geral, com base no modo de transmissão os agentes etiológicos causadores da mastite podem ser classificados grosseiramente como agentes contagiosos ou ambientais. No primeiro grupo, a transmissão ocorre entre os animais, através de materiais contaminados ou através das mãos contaminadas do ordenhador. Por outro lado, os agentes ambientais são transmitidos através do ambiente contaminado [22,24]

Os agentes contagiosos apresentam maior adaptação e estão geralmente associados a casos de infecções subclínicas, casos mais graves que podem levar à perda total de um ou mais quartos mamários. Ao passo que os contaminantes classificados como ambientais, considerados oportunistas são menos adaptados e estão associados a infecções clínicas [25]. Entre os principais agentes contagiosos associados à mastite, observa-se uma maior

prevalência de *S. aureus* e *Streptococcus agalactiae*. Todavia, entre os agentes ambientais mais comuns, encontram-se *Streptococcus uberis*, *Streptococcus dysgalactiae*, *Enterococcus spp.* e *Escherichia coli* [26].

As bactérias *S. aureus*, *S. agalactiae*, *S. uberis* e *E. coli* são normalmente patógenos mais danosos ao úbere que outros agentes etiológicos, por isso são denominados patógenos maiores [2]. Alguns desses são considerados emergentes, como o *S. aureus*, que tem gerado preocupação por ser causador de mastite clínica e subclínica em rebanhos leiteiros, além de poder impactar a saúde humana [18].

Estudos de prevalência, realizados por Mesquita et al. (2019) em 306 rebanhos leiteiros no sul de Minas Gerais, verificaram que o *S. aureus* e *S. agalactiae* foram os patógenos mais frequentes em casos de mastite, com taxas de prevalência de 70,3% e 67%, respectivamente [21]. Dimittimann et al., em estudos de ocorrência de *S. aureus* em cinco laticínios no Brasil, demonstraram que quatro dos cinco laticínios foram positivos para *S. aureus* com prevalência de 0 a 63% entre as amostras. Para os pesquisadores, esses resultados geram preocupações do ponto de vista da saúde pública, devido à alta frequência em rebanhos no Brasil [27].

## 1.2 **Staphylococcus aureus**

O *S. aureus* é uma bactéria do grupo dos cocos gram-positivos, coagulase positivo, que pode se apresentar individualmente, aos pares, em cadeias curtas, ou agrupados como “cachos de uvas”. A denominação *aureus* está relacionada com a pigmentação amarelo-ouro que essa espécie apresenta [28]. Destaca-se por sua diversidade genética, o que lhe garante influenciar o desenvolvimento de doenças infecciosas em hospedeiros humanos e animais [29,30], causando enfermidades que vão desde infecções superficiais de pele e tecidos moles a múltiplas infecções invasivas como bacteremia, pneumonia, osteomielite, etc. [31].

Na mastite, o principal reservatório e fonte da infecção é a glândula mamária, porém, o *S. aureus* pode ser encontrado em outros locais como a pele

e fossas nasais das vacas ou materiais da baia, o que dificulta o controle em rebanhos. [32]. A enfermidade na glândula mamária ocasionada por *S. aureus* compreende infecções de longa duração que podem causar a morte do animal, em parte, por sua alta patogenicidade, que lhe permite alta resistência aos antibióticos [33]. Ademais, a taxa de cura no tratamento de infecções causadas por *S. aureus* é geralmente baixa, cerca de 50%, e apenas se o tratamento iniciar antes do desenvolvimento completo do estado crônico [26].

Em suma, a alta patogenicidade apresentada por *S. aureus* está associada a seu perfil de virulência, que resulta em maior facilidade de multiplicação, penetração nos tecidos mamários, adaptação e maior sobrevivência em diferentes hospedeiros e em diferentes locais do corpo [34].

#### 1.2.1 Tratamento

Programas clássicos de controle de mastite, baseados em estratégias de higiene e antibioticoterapia, recomendam o uso de antibióticos em casos clínicos e subclínicos independente do estado infeccioso da glândula mamária [35]. A terapia com antimicrobianos é a estratégia mais importante no controle e tratamento de mastite por *S. aureus* [36].

Os antibióticos são mais comumente administrados por via intramamária, especialmente para terapia de vacas fora do período de lactação [37] ou administrados por via sistêmica para o tratamento de mastite clínica, geralmente para eliminar o patógeno causador, prevenir o agravamento da doença e para o rápido retorno do leite comercializável das vacas afetadas [2]. As formulações antimicrobianas, como  $\beta$ -lactâmicas (penicilinas, cefaloporinas de terceira e segunda geração), fluoroquinalonas, aminoglicosídeos, tetraciclina, sulfonamidas são as classes de medicamentos mais utilizadas mundialmente para o tratamento de infecções intramamárias, clínicas e subclínicas [35].

No tratamento de mastite subclínica em rebanhos brasileiros, geralmente é adotada a terapia combinada administrada via intramamária com cefalosporina de terceira geração, seguida por antimicrobianos com combinação de tetraciclina aminoglicosídeo e polipeptídeo ou outra combinação de amplo

espectro que seja aprovado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) [2].

A seleção do antimicrobiano e das estratégias na via de administração, é normalmente baseada na identificação do agente infeccioso, nos tipos de sinais apresentados (clínico ou subclínico) e nas características associadas à vaca. No entanto, do ponto de vista das infecções causada por *S. aureus*, a resistência às classes de fármacos de primeira, segunda e terceira linha causa uma grande preocupação à saúde pública, com relevância global para a saúde animal, especificamente para o gado [38].

### 1.2.2 Resistência antimicrobiana

Nas últimas décadas, vários patógenos vêm evoluindo com resistência a classes de fármacos de primeira, segunda e terceira linha, tornando-os quase impossíveis de serem tratados [7]. A resistência antimicrobiana (RAM), adquirida por cepas de *S. aureus*, frequentemente identificada em vários países, incluindo o Brasil, representa a preocupação mais relevante atualmente no controle da mastite bovina assim como infecções humanas [37].

O desenvolvimento de resistência de *S. aureus* a muitos antibióticos surge geralmente por mutações que alteram os sítios de ligações dos fármacos em alvos moleculares [39]. A resistência intrínseca ou natural do *S. aureus*, mostrou aptidão reduzida na presença de ciprofloxacina, daptomicina, gentamicina ou oxaciclina durante a ativação de fatores como *mprF*, *ndh*, *fmtA*, *graR* ou *dltA*. A ativação de fatores de bombas de efluxo é codificada cromossomicamente, como *norA*, que lhes confere resistência ou baixa suscetibilidade à classe de quinolonas [40].

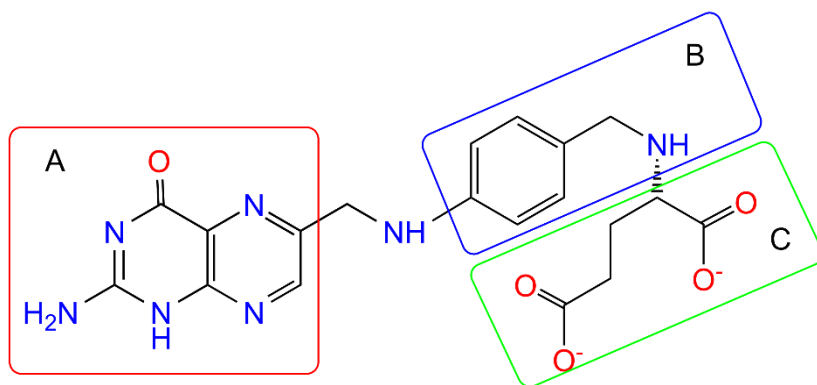
Dados provenientes da Organização Mundial da Saúde (OMS) destacam a resistência do *S. aureus* como uma das principais ameaças de saúde global e coloca o *S. aureus* na lista de prioridade para pesquisa e desenvolvimento de novos antibióticos [18]. Neste sentido, estudos avaliando a prevalência da RAM entre cepas de *S. aureus* são importantes para direcionar estratégias de tratamento e garantir o uso prudente de antimicrobianos [39].

Em geral, cepas de *S. aureus* isoladas de casos de mastite apresentaram alta ocorrência de resistência, principalmente aos antibióticos: ceftarolina, trimetoprima-sufametoxazol [41], meticilina, penicilina e tetraciclina, os quais ainda são os principais medicamentos usados em casos de mastite [42].

### 1.3 DHFR alvo molecular promissor para desenvolver antibacteriano

DHFR é uma enzima presente tanto em células procariontes quanto em eucariontes. O papel da DHFR está relacionado com a via metabólica do folato, que está envolvido em múltiplas vias biossintéticas, incluindo biossíntese de timidilatos (precursor da replicação do DNA) [43]. Os folatos são vitaminas B9 compostas por três subestruturas: anel pterino, ácido para-aminobenzoico (PABA) e o ácido glutâmico (**Figura 1**). Além disso, são encontrados em sua forma oxidada, como ácido fólico, ou como folatos reduzidos, diidrofolato (DHF) ou tetraidrofolato (THF) [44].

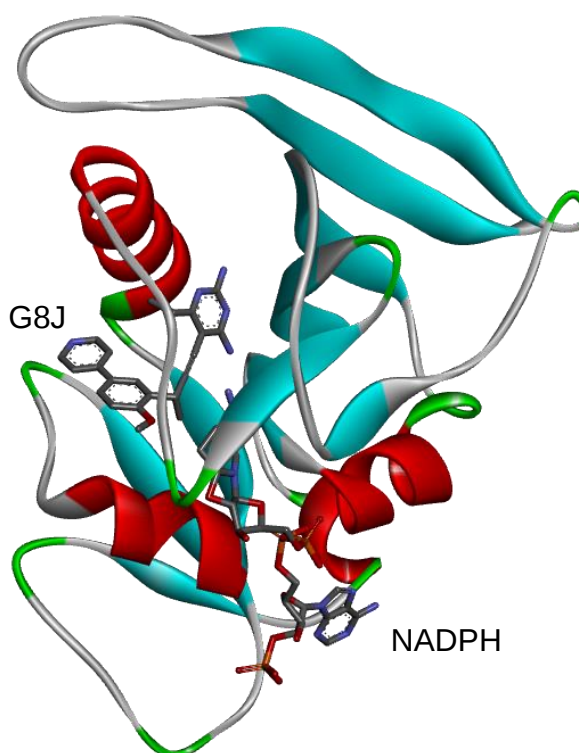
**Figura 1:** Estrutura do folato e suas três subestruturas. (A) pterina; (B) ácido para-aminobenzoico; (C) ácido glutâmico.



Fonte: próprio autor

A via do folato é essencial para a biossíntese do DNA e tem sido explorada de forma farmacológica no tratamento de diversas doenças como câncer, malária, toxoplasmose, artrite reumatoide, infecções bacterianas e fúngicas [44,45]. Os inibidores da enzima DHFR impede que cofatores de DHF se reduza a THF na presença da coenzima NADPH, resultando na morte celular da bactéria. Por este motivo, a DHFR é um alvo promissor para desenvolvimento de fármacos contra *S. aureus* [46]. Na **Figura 2** é possível observar a estrutura da SaDHFR em forma de fita.

**Figura 2:** Estrutura da DHFR do *S. aureus* na representação de fita.

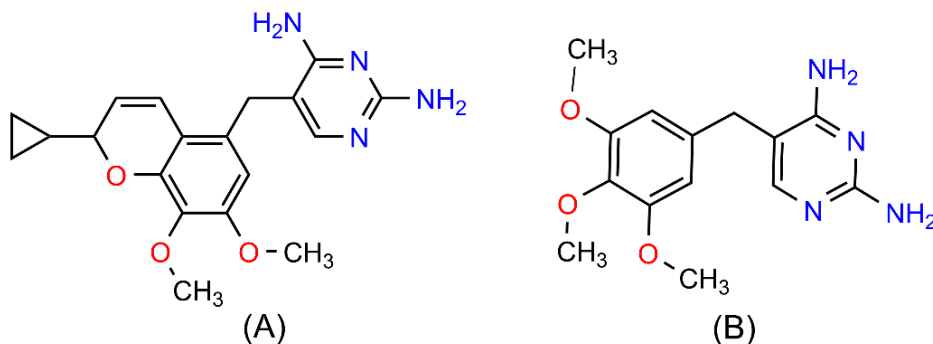


Fonte: Adaptado do arquivo PDB de código 6ND2. Estrutura 3D obtida por Lombardo et al., 2019. A DHFR apresenta subdomínios de ligação à adenosina e de alças [47].

O sítio ativo da DHFR é composto de alfas hélices e folhas betas encahadas. Esta enzima apresenta dois subdomínios: o domínio de ligação ao substrato e o subdomínio adenosina, onde a coenzima NADPH ocupa maior parte. Os inibidores de DHFR se ligam especialmente por ligações hidrofóbicas, ligações de hidrogênio, interações de empilhamento  $\pi$  [18,48].

Os medicamentos que se ligam inibindo a DHFR, considerados antifolatos, podem ser agrupados em duas classes: lipofílicos e clássicos. A trimetoprima (TMP) e iclaprim (**Figura 3**), ambos com atividade antibacteriana, estão entre os inibidores de DHFR lipofílicos, pois contêm um farmacóforo 2,4-diaminopirimidina e se difundem passivamente no espaço citosólico. Enquanto o metotrexato e o pemetrexede, ambos quimioterápicos, são exemplos de antifolatos clássicos, pois apresentam uma porção glutamato [49].

**Figura 3:** Estruturas químicas dos inibidores de DHFR: iclaprim (A), trimetoprima (B).



Fonte: próprio autor

A TMP, uma 2,4-diaminopirimidina aprovada pela FDA e ANVISA, é um antimicrobiano capaz de inibir de forma competitiva o metabolismo dos ácidos nucleicos. Durante o processo de ligação da enzima ao TMP (DHFR-TMP), o grupo aminopirimidina protonado de TMP se liga ao grupo carboxilato da enzima, através de um par de ligação de hidrogênio N-H-O [50]. A trimetoprima atua com mais seletividade pela enzima DHFR de bactéria, quando comparada à enzima humana [51], apresentando uma concentração de inibição contra a DHFR do *S. aureus* de 1,24 nanomolar (nM), enquanto que para enzima humana concentração de inibição é de 19700 nM [52].

O perfil farmacológico favorável do trimetoprima, como boa biodisponibilidade absoluta, logo selecionou microrganismos resistentes, dentre eles o *S. aureus* que apresentou resistência ao TMP através de mutações no sítio ativo da DHFR. A DHFR com mutação (F98Y) dificulta a formação de ligação de hidrogênio desse aminoácido com o grupo amino da trimetoprima, aumentando a energia livre de ligação TMP-DHFR, elevando, consequentemente, a concentração mínima de inibição [53]. O iclaprim, outro inibidor da DHFR de bactérias, desenvolvido para inibir a DHFR do *S. aureus* com a mutação F98Y (SaDHFR<sub>F98Y</sub>), apresenta uma rápida ação bactericida, é mais seletivo que o TMP pela enzima com a mutação F98Y, e apresenta IC<sub>50</sub> de 27 nM. Contudo, mesmo apresentando características *in vitro* e clínicos promissores, o iclaprim apresenta baixa biodisponibilidade oral quando comparado ao TMP [54,55]. Diante disso, torna-se necessário o desenvolvimento de novos inibidores de SaDHFR mais eficientes, com

propriedades farmacocinéticas melhores [49]. Assim, métodos computacionais podem ser empregados para auxiliar na descoberta de novas substâncias com propriedades farmacocinéticas.

#### 1.4 Químioinformática no planejamento de fármacos

A descoberta de novos fármacos é um processo complexo, longo e envolve altos investimentos [56]. As técnicas computacionais, como a triagem robotizada de alto rendimento (HTS, do inglês *high-throughput screening*) e as técnicas de triagem virtual (VS, do inglês *Virtual Screening*) são amplamente utilizadas na busca de novas substâncias biologicamente ativas. Porém a HTS possui alto custo e baixa taxa de acerto quando comparada a VS, usada como uma estratégia para redução de tempo e custos no processo de planejamento de novos medicamentos [57].

O planejamento de fármacos auxiliado por computador (CADD, do inglês *Computer-Assisted Drug Design*) contribui significativamente para Química Medicinal, como análise de dados biológicos e químicos e simulações de eventos moleculares. Neste sentido, as abordagens em químioinformática e modelagem molecular apresentam grande destaque e são geralmente utilizadas com os seguintes fins: 1) filtrar extensas bibliotecas de estruturas químicas, reduzindo-as a subconjuntos com propriedades desejadas; 2) otimizar séries de estruturas químicas em relação às suas propriedades farmacodinâmicas e farmacocinéticas; 3) planejar novas substâncias através de informações já conhecidas sobre o alvo molecular [58].

A químioinformática (do inglês, *Chemoinformatics* ou *Cheminformatics*) ou simplesmente “química *in silico*” é uma ciência interdisciplinar que utiliza metodologias das ciências da computação e da informação para resolver problemas da química. A químioinformática, atualmente, se concentra no *design* de medicamentos, engenharia molecular e química orgânica [59].

De maneira geral, as estratégias em químioinformática, são classificadas em abordagens baseadas em ligantes (LBDD, do inglês *Ligand-Based Drug Design*) e abordagens baseadas em estruturas (SBDD, do inglês *Structure-*

*Based Drug Design*). Entre as abordagens SBDD, destaca-se a docagem molecular, que é geralmente utilizada na predição de complexos ligantes-receptor. E, por outro lado, dentre as ferramentas LBDD, destacam-se os modelos de relações quantitativas entre estrutura e atividade (QSAR, do inglês *Quantitative Structure-Activity Relationships*) [56].

#### 1.4.1 QSAR

O campo da modelagem QSAR foi inicialmente popularizado em 1962 por Corwin Hansch e é uma das ferramentas computacionais mais aplicada na física orgânica e química medicinal [60]. Historicamente, muitos algoritmos matemáticos que englobam diferentes áreas de aplicação em diversas pesquisas, incluindo nanotecnologia, ciências de materiais, biomateriais, informática clínica e planejamento de síntese foram integradas as técnicas modernas de aprendizagem de máquina [61].

Atualmente, a modelagem QSAR é uma técnica de aprendizado de máquina importante para a descoberta de novos medicamentos. Os estudos de QSAR são uma estratégia estabelecida para identificar características estruturais que se correlacionam com a atividade biológica ou outra propriedade alvo. Os modelos QSAR são construídos estabelecendo relações empíricas, lineares e não lineares entre um conjunto de dados, formado por estruturas químicas associadas a uma propriedade alvo, com os valores de descritores moleculares, calculados a partir das estruturas químicas [61].

Os descritores usualmente aplicados em modelos QSAR são o resultado de cálculos matemáticos e lógicos que transformam as informações estruturais de uma molécula em um número útil para o desenvolvimento de um modelo. Os descritores podem ser classificados de acordo com a dimensionalidade da representação da molécula usada no cálculo, como unidimensional (1D), bidimensional (2D) e tridimensional (3D). Este último, depende das conformações tridimensionais das moléculas [24].

Outra classificação está relacionada à natureza dos descritores, que podem ser: constitucionais (ex. número de anéis, número de átomos e massa molar); índices topológicos (ex. índice de Zagreb); geométricos (ex. valor do

componente x quadrupolo / ponderado pela polarizabilidade); descritores de distribuição radial (RDF, o inglês *Radial Distribution Function*), entre outros. O número e tipos de descritores dependem da qualidade do conjunto de dados estudados e do algoritmo usado para o cálculo [24,62]. Os descritores combinados são aplicados em diferentes algoritmos matemáticos de máquina de aprendizagem que os correlacionam com uma variável independente [63].

Entre os algoritmos popularmente utilizados para modelar dados estão a Árvore de Decisão, Redes Neurais Artificiais (ANN - *Artificial Neural Networks*), Mínimos Quadrados Parciais, k-vizinho-mais-próximo (KNN - *K-Nearest Neighbors*), Regressão Linear Múltipla (MLR - *Multiple Linear Regression*), Análise Discriminante e Máquinas de Vetores de Suporte (SVM - *Support Vector Machines*). Os quais são suportados por abordagens da estatística e aprendizado de máquina [57,64].

O desempenho de cada modelo passa por avaliações estatísticas de validação interna e externa, objetivando obter modelos estatisticamente apropriados e preditivos e evitar uma relação ao acaso. No âmbito das avaliações estatísticas utilizadas estão, entre outros parâmetros, o coeficiente de determinação (treino:  $R^2$ , validação cruzada:  $Q^2$  e validação externa/teste externo:  $P^2$ ), Erro Absoluto Médio (MAE do inglês *mean absolute error*) e da Raiz do Erro Quadrático Médio (RMSE do inglês *root mean squared error*). O  $R^2$  é a proporção da variação da variável dependente que é previsível a partir das variáveis independentes, ou seja, quantifica o quanto o modelo é capaz de prever os dados experimentais e seu cálculo é feito a partir da equação (1) [65]. Já o MAE representado pela equação (2) é uma média aritmética dos erros absolutos entre as observações pareadas aos valores previstos, média absoluta dos resíduos e apresenta mesma unidade de medida da variável dependente. Por fim, o RMSE, calculado pela equação 3, representa a raiz quadrada da média dos erros quadrados, que está relacionado à comparação da performance em dados externos e internos [66,67]. Os argumentos usados nas equações são:  $y_{ipred}$  é o valor predito correspondente;  $\overline{y_i}$  é a média correspondente e  $n$  o número de amostras utilizadas no modelo.

$$R^2 = 1 - \frac{\sum (y_{iexp} - y_{ipred})^2}{\sum (y_{iexp} - \bar{y}_i)^2} \quad (1)$$

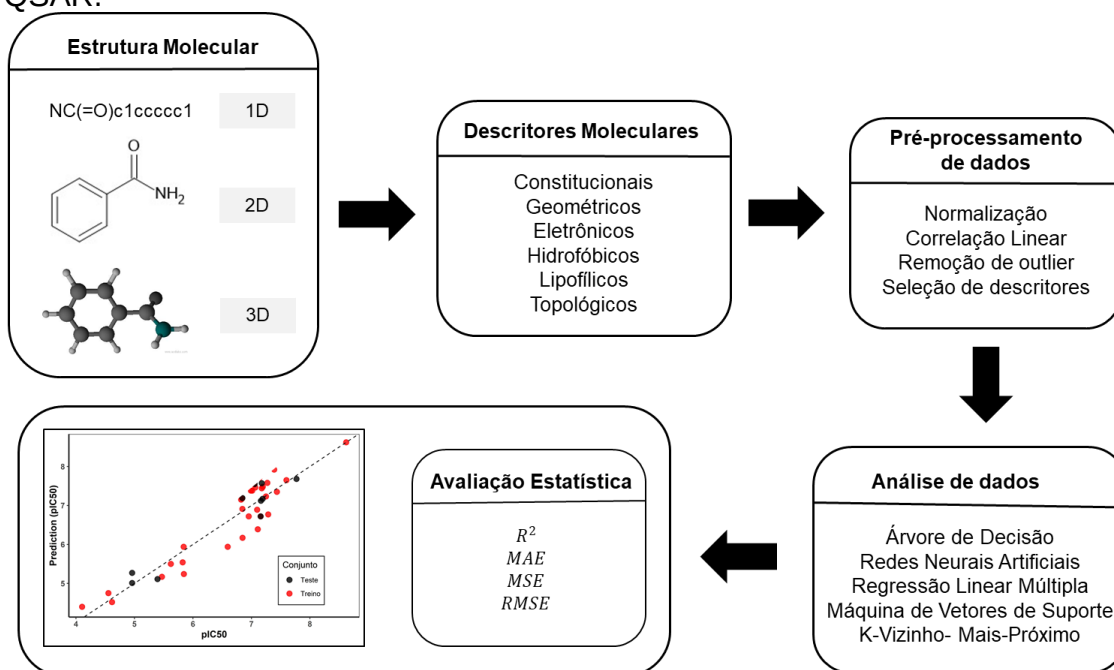
$$MAE = \left| \frac{\sum (y_{iexp} - y_{ipred})}{n} \right| \quad (2)$$

$$RMSE = \sqrt{\frac{\sum (y_{iexp} - y_{ipred})^2}{n}} \quad (3)$$

Além disso, de acordo com a Organização para a Cooperação e Desenvolvimento Econômico (OECD, do inglês *Organization for Economic Cooperation and Development*), cinco princípios básicos são considerados para a geração e validação de modelos: (i) um ponto final definido; (ii) algoritmo inequívoco; (iii) domínio de aplicabilidade definido; (iv) apresentar robustez e previsibilidade; (v) realizar a interpretação mecanística dos modelos, quando possível [68].

Nesse contexto, os modelos desenvolvidos e validados fornecem informações importantes para prever novas moléculas biologicamente ativas, que podem posteriormente ser testadas experimentalmente [69]. Assim, claramente, a modelagem QSAR é uma abordagem extremamente útil e modelos validados podem fornecer orientações valiosas para muitas áreas de pesquisa, em especial para o direcionamento de novos medicamentos [61]. Na **Figura 4** é mostrado um esquema resumido do processo de desenvolvimento de um modelo QSAR.

**Figura 4:** Esquema simplificado do processo de desenvolvimento do modelo QSAR.



Fonte: DAMALE et al., 2014 (Adaptado) [70]

#### 1.4.2 Docking Molecular

Em termos simples, *docking* molecular, acoplamento molecular ou ancoragem molecular é uma técnica de modelagem SBDD usada para posicionar uma estrutura 3D (pequena molécula ligante ou uma proteína) em uma estrutura receptora (proteína, DNA). Esse método é útil no desenvolvimento de medicamentos e possibilita o estudo das interações intermoleculares em sistemas complexos como, proteína-proteína, ligante-proteína e proteína - DNA [71].

As técnicas de acoplamento abordam três objetivos principais: triagem virtual, acoplamento e estimativa de afinidade de ligação. As ferramentas aplicadas para acoplamento incluem algoritmos que visam a busca da melhor orientação (pose que o ligante pode assumir dentro do sítio ativo) e *scores* (pontuação de avaliação das poses do ligantes) [72].

Uma preparação da proteína e do ligante é necessária antes do acoplamento. Essa preparação ocorre separadamente onde é obtida a conformação tridimensional do ligante, ao receptor é adicionado hidrogênios e

as moléculas de água removidas se consideradas interferentes. Além disso, um protocolo de acoplamento deve ser validado para tornar os resultados confiáveis. Nesse caso um encaixe focado é recomendado. O encaixe focado, trata-se de um método de *redocking* onde o ligante cocrystalizado com a enzima é separado e reencaixado no sítio ativo assumindo diferentes conformações [73].

O estudo realizado por Li et al. 2011, aplicou a técnica baseada em estrutura para descobrir inibidores seletivos de SaDHFR utilizando informações das estruturas cocrystalizadas. O derivado 7-aril-2,4-diaminoquinazolina obteve o maior sucesso com base nessa abordagem e demonstrou interações de ancoragem com resíduos hidrofóbicos no sítio SaDHFR [52]. Por outro lado, projetos baseados em ligantes também têm sido direcionados. El-Gazzar et al., 2017 aplicou as técnicas baseadas em ligantes para explorar características específicas de um ligante ou modelo molecular. Esse trabalho, a partir das características dos inibidores ativos de quinozolina, identificou análogos ativos de SaDHFR [74].

Estudos combinados baseados em estruturas e ligantes também estão sendo aplicados para descobrir inibidores de DHFR de *S. aureus*. Com base na estrutura cristalina do SaDHFR, Kobayashi et al., 2014, identificaram um composto líder que foi usado como entrada em uma pesquisa de similaridade de uma biblioteca e vários análogos foram identificados como inibidores de DHFR [75]. Matos et al., 2022 também aplicaram as abordagens baseadas em estrutura combinada com as abordagens baseadas em ligante. A ancoragem molecular resultados que sugerem a importância das interações hidrofóbicas entre ligante e os resíduos Ile51, Phe93, Leu55, Val32 e Leu29 [76].

Sendo assim, devido a ineficácia dos medicamentos disponíveis no mercado para o tratamento de infecções causadas por espécies de bactérias há uma urgência na busca de novos inibidores. Além disso, sabendo que a DHFR é um alvo biológico promissor no planejamento de novos medicamentos e que substâncias químicas da classe de alcaloides, apresentam eficácia em estudos *in silico* para descobrir inibidores seletivos de SaDHFR [8].

## 2 JUSTIFICATIVA

Considerando: (i) a ameaça causada por patógenos resistentes, como *S. aureus* tanto para a saúde humana quanto para a saúde animal; (ii) o *S. aureus* ser o principal agente causador de mastite grave em rebanhos brasileiros e em todo o mundo; (iii) os medicamentos disponíveis no mercado já apresentarem ineficácia na maioria das cepas de *S. aureus*; (iv) a urgência de desenvolver novos inibidores para o tratamento de doenças infecciosas causadas por patógeno resistentes, destaca-se que:

- A DHFR é promissora como alvo para descoberta de fármacos no tratamento de doenças infecciosas causadas por patógenos resistentes [9];
- As abordagens auxiliadas por computador no planejamento de fármacos, como abordagens da quimioinformática aceleram a descoberta de novas substâncias com atividade biológica [24,58];
- Os alcaloides representam uma classe de estrutura química com aplicações terapêuticas importantes e já forneceram muitos medicamentos para o tratamento de diversas doenças [7,14,77];
- Estudos *in silico* combinados já demonstraram eficácia na identificação de substâncias ativas contra *S. aureus* e patógenos resistentes [8];
- Estudos de *docking* molecular mostraram que a enzima SaDHFR interage por ligações de hidrogênios, ligações hidrofóbicas e ligações de empilhamento  $\pi$ - $\pi$  [48].

Assim, os resultados encontrados no presente trabalho poderão direcionar o planejamento de fármacos no tratamento de doenças infecciosas causadas em humanos e animais. Além disso, vale ressaltar que, até onde se sabe, este trabalho é o único a unir um banco de diferentes classes de alcaloides para selecionar novos inibidores de SaDHFR.

### 3 OBJETIVOS

#### 3.1 Objetivo Geral

Identificar alcaloides potencialmente inibidores de *S. aureus* e da enzima SaDHFR.

#### 3.2 Objetivos Específicos

- Selecionar um banco de dados de estruturas químicas com atividade contra *S. aureus* e na enzima SaDHFR;
- Construir um banco de dados com estruturas químicas de alcaloides;
- Calcular descritores bidimensionais (2D) e tridimensionais (3D);
- Selecionar, a partir dos descritores calculados, descritores mais correlacionados com a atividade de inibição do *S. aureus* e da SaDHFR;
- Desenvolver e validar modelos QSAR para previsão de novas substâncias que possam inibir o *S. aureus* e a SaDHFR;
- Realizar triagem virtual no banco de alcaloides, a partir dos modelos previamente desenvolvidos e validados para descoberta de possíveis alcaloides ativos;
- Realizar estudo *in silico* combinando abordagens baseadas em ligante e baseadas em estrutura para seleção de alcaloides com potencial atividade contra SaDHFR;
- Prever o modo de ligação dos alcaloides no sítio ativo da SaDHFR através de *docking* molecular.

## 4 MATERIAL E MÉTODOS

### 4.1 Computador e Softwares

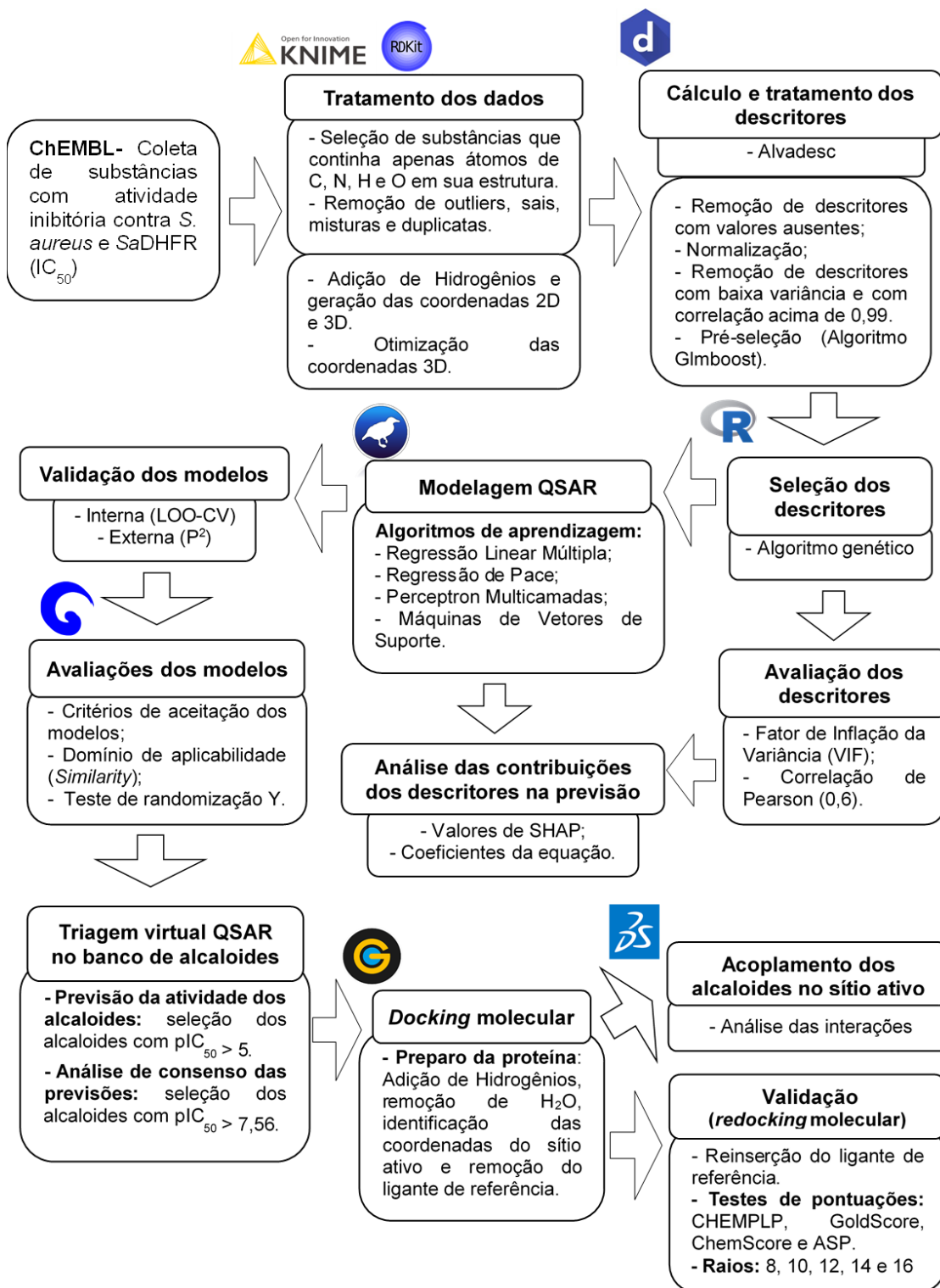
Os cálculos foram executados com o auxílio de um computador Intel® Core i3 com 4GB de memória RAM com sistema operacional *Microsoft Windows* 10. Os softwares utilizados foram:

- *Knime Analytcs Platform*, versão 4.4.2 (KNIME AG, Alemanha), para desenvolver e validar os modelos;
- *BIOVIA Discovery Studio*, versão 21.1.0.0 (Dassault Systemes, França) para leitura dos dados;
- *MOPAC2016 (Molecular Orbital PACkage)*, versão 22.097L, para calcular a conformação tridimensional;
- *Alvadesc*, versão 2.0 (Alvascience, Itália), para calcular os descritores (<https://www.alvascience.com/>);
- *R*, versão 4.2.3 (Universidade de Auckland, Nova Zelândia), para mineração dos dados;
- *ChemSketch*, versão 2.1, (ACD/Labs, Canadá) para desenhar as estruturas;
- *RDKit - Open-Source Cheminformatics Software*, (<https://www.rdkit.org>) para preparar as estruturas;
- *WEKA - Waikato Environment for Knowledge Analysis*, versão 4.2.0 (Universidade de Waikato, Nova Zelândia) integrados ao Knime para desenvolver os modelos;
- *GOLD - Genetic Optimisation for Ligand Docking*, versão 2022.3.0 (Cambridge Crystallographic Data Centre, Reino Unido) foi utilizado para realizar o *docking* molecular.

### 4.2 Métodos

Um esquema geral das etapas empregadas neste trabalho pode ser visualizado na **Figura 5**.

**Figura 5:** Esquema geral das etapas empregadas na identificação de alcaloides inibidores de *S. aureus*.



Fonte: próprio autor.

#### 4.2.1 Modelagem quantitativa de estrutura-atividade para *S. aureus* e SaDHFR

##### 4.2.1.1 Conjunto de dados de entrada

Para os estudos de modelagem de QSAR foram selecionados dois bancos de dados da base ChEMBL [78]. O primeiro formado por estruturas químicas testadas contra o *S. aureus* e o segundo com estruturas químicas testadas na enzima SaDHFR.

O conjunto de estruturas químicas coletado a partir do ChEMBL (AID: 352) consiste em 64.388 estruturas, das quais foram coletadas 2.676 estruturas com valores médios de concentração de 50% de inibição ( $IC_{50}$ ) para *S. aureus*, este conjunto de estruturas foi nomeado de “SaDB”. O segundo conjunto de estruturas químicas foi obtido da base ChEMBL (AID: 1681620) o qual possui 208 estruturas químicas, das quais foram obtidas 117 com valores de  $IC_{50}$  para SaDHFR, que será chamado de “SadhfrDB”. Ambos os conjuntos de estruturas foram utilizados para construção de modelos de regressão.

##### 4.2.1.2 Preparo do Conjunto de Dados

A modelagem QSAR é baseada na premissa de que substâncias estruturalmente semelhantes exibem efeitos biológicos próximos [62]. Neste sentido, para obter um conjunto de dados estruturalmente similar aos alcaloides, foram selecionados para desenvolver os modelos apenas substâncias que apresentavam em sua estrutura molecular átomos de C, H, N e O.

O sucesso de qualquer estudo em quimioinformática, especialmente para a modelagem QSAR parte de dados cuidadosamente preparados [79]. Assim, ambos os conjuntos de dados passaram por curadoria. Nessa etapa, foram removidas as substâncias que não se adequavam às abordagens da quimioinformática, como estruturas e subestruturas de substâncias inorgânicas, sais e misturas [80]. Além disso, todos os valores de atividade ( $IC_{50}$ ) foram padronizados para a unidade molar (M), as estruturas duplicadas e que apresentavam valores atípicos de  $IC_{50}$  da maior parte do conjunto (*outliers*) foram descartadas dos conjuntos SaDB e SadhfrDB. Os valores atípicos de  $IC_{50}$  foram selecionados aplicando o intervalo interquartil (IQR). Tais valores foram

removidos antes de prosseguir com a construção dos modelos, visto que poderiam levar à instabilidade dos modelos [81]. Um valor de  $IC_{50}$  foi considerado como atípico quando estava fora do intervalo:

$$R = [Q_1 - k(IQR), Q_3 + k(IQR)] \quad (4)$$

Onde,  $Q_1$  e  $Q_3$  são o primeiro e terceiro quantil;  $k = 1,5$ ; e,  $IQR = Q_3 - Q_1$ .

Após tratamento, restaram 405 estruturas no conjunto *SaDB* e 47 estruturas no conjunto *SadhfrDB*. As concentrações da atividade foram expressas como logaritmos ( $-\log IC_{50}$  representado por  $pIC_{50}$ ). A todas as estruturas químicas restantes nos conjuntos foram adicionados hidrogênios explícitos, as coordenadas tridimensionais foram geradas baseadas nas distâncias, pré-otimizadas usando o campo de força MMFF94s com os pacotes do RDKit [82] contidos no Knime e otimizadas utilizando o método semiempírico PM7, no software MOPAC 2016 [83]. O método semiempírico é baseado na mecânica quântica e objetiva capturar propriedades da mecânica quântica com um custo computacional reduzido [84].

#### 4.2.1.3 Cálculo dos descritores moleculares

As estruturas otimizadas foram usadas para calcular os descritores no AlvaDesc [85]. Este software realiza cálculos de uma variedade de descritores 0D, 1D, 2D e 3D. Para os dois conjuntos de dados foram calculados descritores 2D e 3D no conjunto *SadhfrDB* os descritores foram mantidos separados objetivando construir modelos com os descritores 2D e 3D separadamente. Construir modelos separados apresenta vantagens, pois demanda custo computacional menor e podem ser realizados modelos de consenso que garantem modelos mais confiáveis [86].

#### 4.2.1.4 Pré-tratamento dos descritores moleculares

A seleção de descritores em um estudo de QSAR é a etapa mais importante, pois afeta diretamente a qualidade dos modelos. Desta forma, ao lidar com inúmeros descritores moleculares que podem representar variáveis redundantes, é importante realizar inicialmente uma etapa de pré-tratamento antes da etapa de seleção do subconjunto mais importante [62]. Os descritores

redundantes tendem a fazer com que o sistema de seleção concentre a atenção nas características de poucas estruturas e desconsidere a generalização para explicar o conjunto [87].

Nesse sentido, foi realizado um pré-tratamento nos descritores calculados. Inicialmente, os descritores com valores ausentes foram excluídos. Em seguida, os descritores normalizados com valores de 0 a 1 (equação 5) passaram por filtro de baixa variância e aqueles que apresentaram variância igual ou próxima a zero foram excluídos.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

Onde  $X'$  é o descritor normalizado,  $X$  é o descritor original,  $X_{max}$  é o valor máximo do descritor original e  $X_{min}$  é o valor mínimo do descritor original.

Além disso, os descritores com correlação linear entre si acima de 0,99 também foram descartados da matriz de dados, pois são considerados descritores redundantes [88]. Em seguida, os dados foram divididos randomicamente em conjunto de treino (corresponde a 80% das estruturas) e conjunto de teste (corresponde a 20% das estruturas).

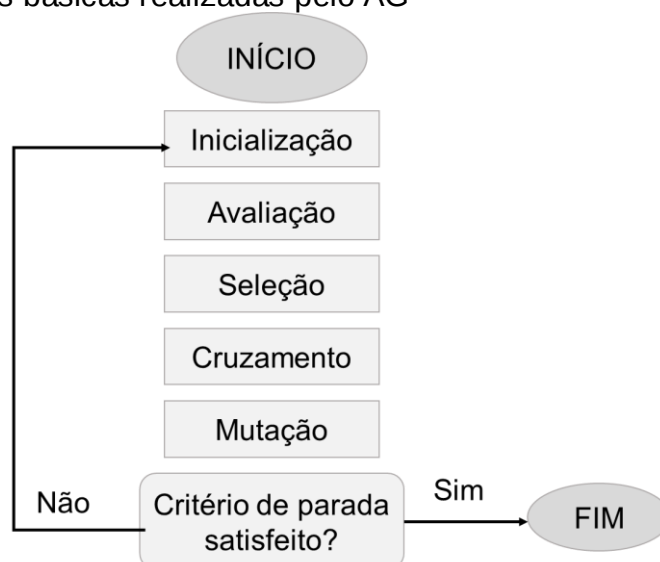
Uma avaliação da importância das contribuições dos descritores foi executada no conjunto SaDB. Para isso, o algoritmo Glmboost do pacote mboost foi utilizado com a função varImp contida no pacote Caret, presente na linguagem R e integrada ao Knime. A avaliação da importância dos descritores foi feita no conjunto de teste [89].

#### 4.2.1.5 Seleção dos descritores moleculares

O conjunto final obtido após o pré-tratamento foi usado como variável de entrada para selecionar o subconjunto de descritores mais relevante. Na literatura, o método de seleção por algoritmo genético (AG) é reconhecido por estudos como uma ferramenta muito eficaz em análises de dados, principalmente para reduzir a dimensionalidade das variáveis em modelos de regressão [90,91].

O AG é uma ferramenta que imita a evolução da vida na Terra (cruzamento, seleção natural e mutação genética), investiga vários subconjuntos simultaneamente, explorando diferentes regiões no espaço vetorial definido pelos descritores, apresentando os que melhor evoluíram no processo e aqueles de maior importância para descrever a atividade biológica [92,93]. O AG foi utilizado para selecionar as combinações de descritores mais relevantes. Para isso, variou-se as seguintes configurações: o número de gerações de 500 até 1300; o número de população de 100 a 300; e o número de variáveis de 15 a 20 no conjunto de dados *SaDB* e 4 a 8 no conjunto de dados *SadhfrDB*. As seleções resultantes para o conjunto *SadhfrDB* serão chamadas de *SadhfrDB-2D* (seleção de descritores 2D) e *SadhfrDB-3D* (seleção de descritores 3D). A **Figura 6** mostra as etapas básicas realizadas pelo AG.

**Figura 6:** Etapas básicas realizadas pelo AG



Fonte: adaptado de Crisan, 2010

Dentre os subconjuntos selecionados, foram considerados para modelagem QSAR os subconjuntos que n o apresentaram descritores colineares, ou seja, descritores com correla  o entre si, menor do que 0,6; pois os descritores selecionados devem apresentar a menor intercorrela  o tanto quanto poss vel [94]. O fator de infla  o da vari ncia (VIF) dos descritores tamb m foi verificado. Um valor estimado de VIF igual a 1 significa que n o existe correla  o entre os descritores; todavia, se o valor estiver entre 1-5 a correla  o   aceit vel; por fim, se for maior que 10 existe multicolinearidade e o modelo  

rejeitado [95]. O fator de inflação da variância do modelo (VIF) é calculado pela equação 6:

$$VIF = \frac{1}{1 - R_j^2} \quad (6)$$

Onde  $R_j^2$  é o coeficiente de determinação do descritor j sobre os demais descritores.

#### 4.2.1.6 Construção dos modelos

Nessa etapa, objetivou-se desenvolver modelos lineares e não lineares, usando quatro diferentes algoritmos de aprendizagem de máquina *pace regression* (PR), regressão linear (ou *linear regression* - LR), ANN do tipo *multilayer perceptron* (MLP) e *support vector machine* (SVM). Para o conjunto de modelagem SadhfrDB, foram construídos modelos separados para as seleções SadhfrDB-2D e SadhfrDB-3D.

A melhor seleção de descritores obtida pelo algoritmo AG para cada conjunto de descritores, foi aplicada nos métodos de aprendizagem de máquina, PR, LR, MLP e SVM. Os dois primeiros métodos tratam-se de métodos lineares, enquanto os dois últimos são métodos de relação não linear. Os 4 algoritmos aplicados para a modelagem QSAR correspondem a algoritmos contidos no software WEKA [96] integrados no Knime.

#### 4.2.1.7 Validação dos modelos

Os modelos desenvolvidos foram avaliados quanto a sua robustez e capacidade preditiva considerando os parâmetros estatísticos obtidos no modelo ajustado (grupo treino),  $R^2$ , na validação interna,  $Q^2$ , na validação externa (grupo teste),  $P^2$ , e os valores de MAE e RMSE.

##### 4.2.1.7.1 Validação interna leave-one-out

A validação interna baseada em validação cruzada *leave-one-out* (deixe um de fora) foi usada para avaliar a robustez e estabilidade dos modelos construídos. Neste método, é removido do conjunto de treino uma estrutura química e apenas com as estruturas restantes  $Q^2$  é calculado, em seguida, a

estrutura removida é testada. O mesmo ocorre até que todas as estruturas do treino sejam testadas e o  $Q^2$  calculado. No entanto, a validação de um modelo QSAR baseada apenas em avaliações de mérito interno pode ser enganosa [97].

#### 4.2.1.7.2 Validação externa

Um modelo QSAR, além de ser estável e robusto (avaliados por validação cruzada interna), também deve ser confiável em previsões de grupos de estruturas químicas que não fizeram parte do grupo treino. Assim, a fim de avaliar a real capacidade preditiva dos modelos, os conjuntos de teste (com 20% das estruturas) foram usados para avaliar a capacidade preditiva dos modelos (validação externa,  $P^2$ ). A qual deve apresentar um valor maior ou igual a 0,6 [94,98].

#### 4.2.1.8 Critério de aceitação e previsibilidade dos modelos

Adicionalmente, foi realizada uma avaliação quanto ao critério de aceitação e previsibilidade dos modelos desenvolvidos, com o pacote Enalos. Os parâmetros avaliados por este pacote determinam que um modelo contínuo e verdadeiro deve atender aos seguintes critérios estatísticos:

$$\begin{aligned}
 R^2 &> 0,6 \\
 R_{VCext}^2 &> 0,5 \\
 \frac{(R^2 - R_o^2)}{R^2} &\leq 0,1 \quad \text{ou} \quad \frac{(R^2 - R_o'^2)}{R^2} \leq 0,1 \\
 0,85 \leq k &\leq 1,15 \quad \text{ou} \quad 0,85 \leq k' \leq 1,15
 \end{aligned}$$

Onde:  $R^2$ : Coeficiente de correlação entre as atividades previstas e reais;  $R_{VCext}^2$ : Validação cruzada externa;  $R_o^2$ : Coeficiente de determinação: atividades previstas versus reais;  $R_o'^2$ : Coeficiente de determinação: atividades reais versus atividades previstas;  $k'$  declive das linhas de regressão de atividades previstas versus reais através da origem; e,  $k$ : declive das linhas de regressão de atividades reais versus previstas através da origem [99,100].

Os modelos também foram validados pelo teste de randomização Y. O teste de aleatorização de Y é um método de validação poderoso que detecta e quantifica uma correlação ao acaso entre a atividade e os descritores. Como

requisito mínimo, um modelo QSAR útil deve descrever os dados fornecidos melhor do que o acaso [101,102].

#### 4.2.1.9 Domínio de aplicabilidade (DA)

Ainda que um modelo QSAR seja validado e considerado robusto, não se pode esperar que ele preveja com confiança para todo o espaço químico das estruturas. O espaço de confiança é definido pelos descritores utilizados para desenvolver os modelos [87]. Assim, os subconjuntos de descritores selecionados para construir os modelos foram avaliados para definir o domínio de aplicabilidade (DA).

Essa etapa foi realizada conforme Zhang et al. (2006), que utiliza medidas de similaridade para estabelecer o DA do modelo a partir das distâncias euclidianas entre todas as estruturas químicas do conjunto de treinamento e as estruturas químicas do conjunto de teste. A distância é comparada a um limite de domínio de aplicabilidade predefinido (APD) [103]. O APD é calculado da seguinte forma:

$$APD = 'd' + Z\sigma \quad (7)$$

Onde: ' $d$ ' é a distância euclidiana;  $Z$  é um valor de corte empírico (padrão 0,5); e,  $\sigma$  é correspondente ao desvio padrão das distâncias [99,100].

#### 4.2.2 Triagem virtual

##### 4.2.2.1 Banco de estruturas de Alcaloides

Este banco foi inicialmente desenvolvido por Paixão (2022) [104] que coletou 820 alcaloides, e no presente trabalho adicionou-se ao banco 380 novos alcaloides presentes em periódicos e na base de dados SciFinder [105]. As estruturas coletadas foram desenhadas utilizando o programa ChemSketch e também validadas no SciFinder.

O banco de dados de alcaloides possui um total de 1200 estruturas químicas de diferentes classes de alcaloides, entre eles: alcaloides quinolínicos, aporfínicos, indólicos, oxindólicos e isoquinolínicos. Neste banco foram realizado as mesmas etapas de tratamento aplicadas nos conjuntos de modelagem.

#### 4.2.2.2 Etapas da triagem

Os melhores modelos QSAR gerados e validados foram utilizados para triar alcaloides promissores em inibir o *S. aureus* e a enzima alvo, SaDHFR. Inicialmente o melhor modelo QSAR do conjunto SaDB foi utilizado para realizar a triagem virtual, onde foram selecionados alcaloides com  $pIC_{50} \geq 5$ .

Em seguida os alcaloides filtrados na triagem virtual com ( $pIC_{50} \geq 5$ ), foram utilizados como entrada para triar os alcaloides mais ativos pelos melhores modelos de cada subconjunto, SadhfrDB-2D e SadhfrDB-3D, ou seja, foram selecionados dois modelos.

Por fim, a seleção dos alcaloides mais ativos foi realizada a partir de um consenso da triagem dos modelos 2D-QSAR e 3D-QSAR, que seguiu com a média simples das atividades previstas dos alcaloides em cada modelo.

#### 4.2.3 Docking Molecular

A proteína alvo de *S. aureus* DHFR foi obtida na plataforma *Protein Data Bank* (PDB) [106] de código 6ND2, complexada com 6-ethyl-5-(3S)-3-[2-methoxy-5-(pyridin-4-yl)phenyl]but-1-yn-1-yl}pyrimidine-2,4-diamine. A proteína foi importada para o programa GOLD e preparada da seguinte forma: (i) foram adicionados átomos de hidrogênio; (ii) foram removidas as moléculas de água; (iii) foi extraído o ligante de referência cristalizado e definidas as coordenadas do sítio. As moléculas de alcaloides selecionadas como potencialmente ativas pelo consenso da triagem foram submetidas a acoplamento molecular (ligante-enzima) no sítio ativo.

Para selecionar os parâmetros do *docking* foi realizado uma validação por *redocking* molecular. Para o *redocking*, após definir o sítio ativo, o ligante extraído foi novamente acoplado no sítio e foram testadas as funções de pontuação CHEMPLP, GoldScore, ChemScore e ASP. Em cada uma das funções de pontuação foram testados diferentes valores de raio: 8, 10, 12, 14 e 16. Os valores de desvio médio padrão (RMSD) obtidos para cada teste foram avaliados. O protocolo de configurações que forneceu o menor valor de RMSD foi selecionado para realizar o acoplamento dos alcaloides ativos.

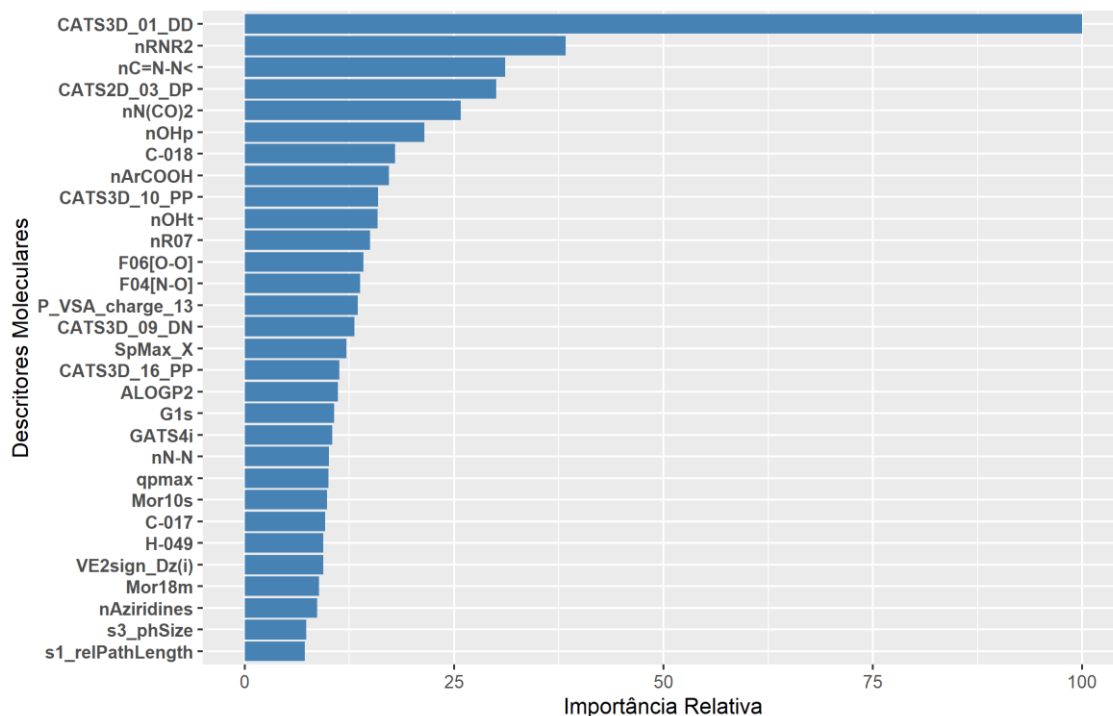
## 5 RESULTADOS E DISCUSSÃO

### 5.1 Modelagem QSAR para identificação de alcaloides inibidores de *S. aureus*

#### 5.1.1 Cálculo e seleção dos descritores

Um total de 5635 descritores moleculares foram calculados para o conjunto SaDB. Na etapa de pré-tratamento, foram excluídos 54 descritores com valores ausentes, 1988 descritores com variância baixa no conjunto de dados e 2321 descritores intercorrelacionados (correlação de Pearson acima de 0,99). Após a etapa de pré-tratamento restaram 1365 descritores, que foram aplicados ao algoritmo Gbmboost para avaliação da importância. Nesta etapa, 60 descritores que apresentavam importância relativa diferente de zero foram selecionados e utilizados como entrada na seleção com por AG. Na **Figura 7**, estão os 30 descritores mais importantes selecionados pelo Gbmboost para o conjunto SaDB.

**Figura 7:** Importância relativa dos descritores selecionados pelo Gbmboost.



As aplicações do método de classificação dos descritores apresentam algumas vantagens como reduzir o custo computacional, aumentar a interpretabilidade dos modelos e reduzir o risco de *overfitting*, pois reduz a

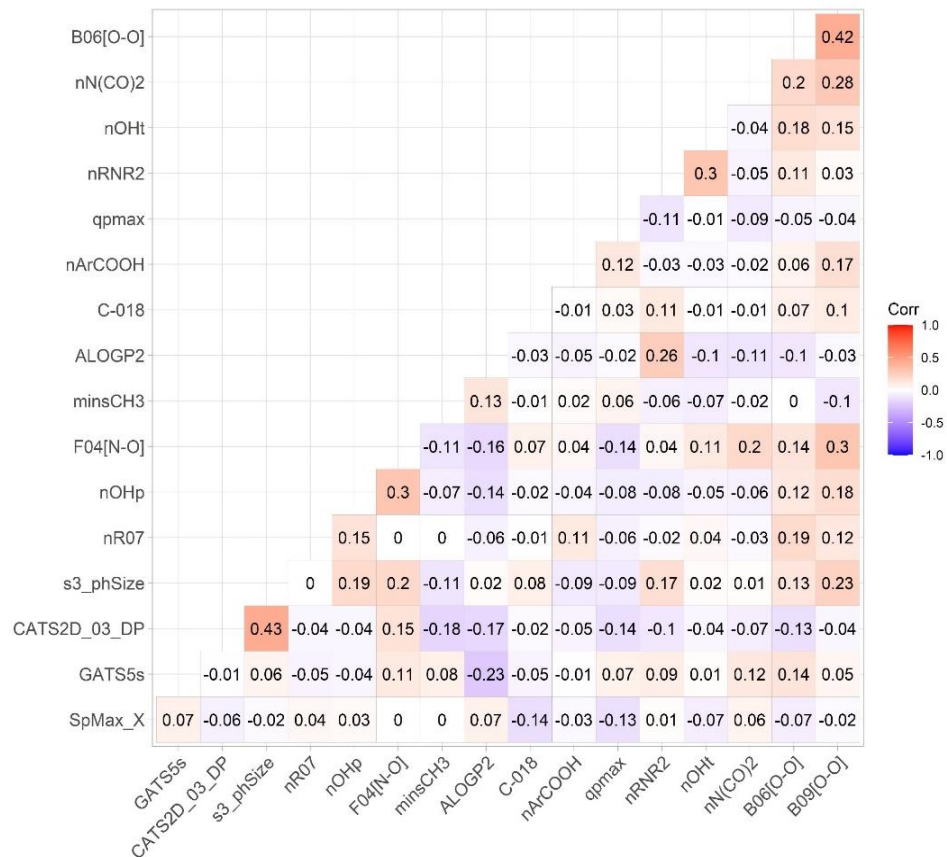
chance de descritores irrelevantes serem selecionados [107]. Um estudo realizado por Du et al. 2022 demonstrou que modelos criados com uma seleção de variáveis apresentaram desempenhos melhores [108].

Para o conjunto SdhfrDB foram calculados 4154 descritores bidimensionais e 1486 tridimensionais. No pré-tratamento do conjunto SdhfrDB-2D foram excluídos 12 descritores com valores ausentes, 2123 descritores com baixa variância no conjunto de dados e 1522 descritores intercorrelacionados. No conjunto SdhfrDB-3D foram excluídos 52 descritores com valores faltantes para algumas moléculas, 1172 descritores com variância limitada e 106 com correlação acima de 0,99. Após o pré-tratamento restaram no conjunto SdhfrDB-2D 497 descritores e no conjunto SdhfrDB-3D 159 descritores.

Ao lidar com um grande número de descritores, a etapa de pré-tratamento dos descritores é fundamental para retirar variáveis redundantes, que podem afetar a qualidade dos modelos [62]. Além disso, as etapas realizadas no pré-tratamento aumentam as chances de identificar um conjunto menor, menos correlacionado entre si e mais correlacionado com a variável resposta e podem fornecer modelos de forma mais eficiente e com custo computacional menor [109].

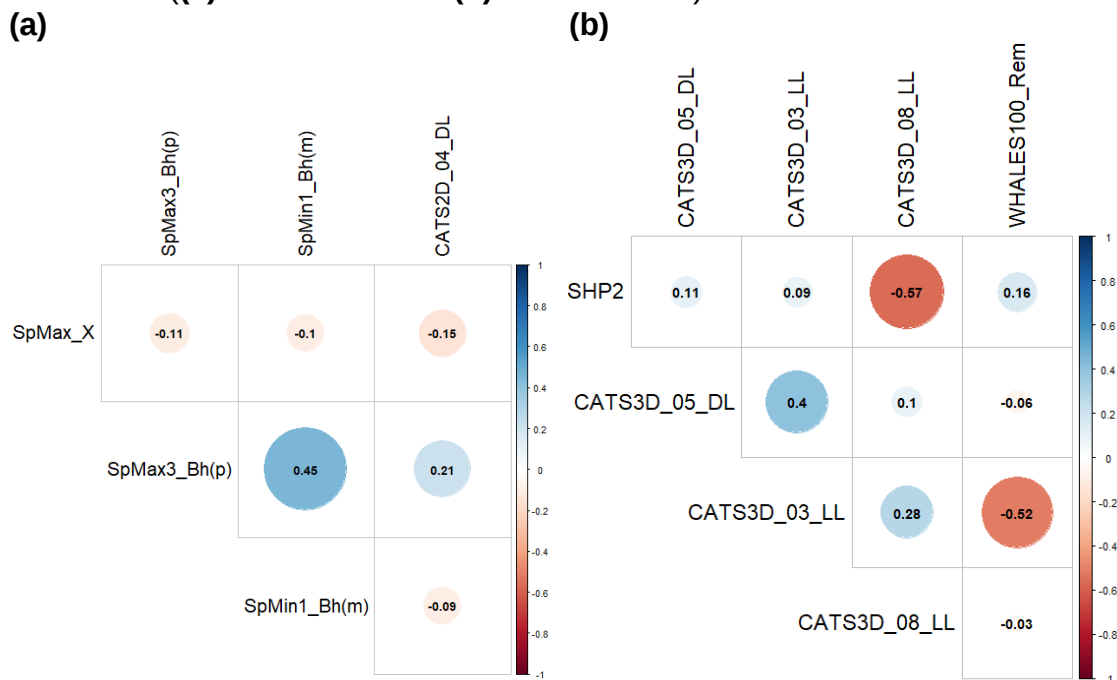
Ao final, para desenvolver os modelos, os conjuntos de descritores pré-tratados foram aplicados como entrada no AG, onde foi selecionado um subconjunto de 17 descritores do conjunto SaDB, um subconjunto de 4 descritores do conjunto SdhfrDB-2D e 5 descritores do conjunto SdhfrDB-3D. As seleções foram avaliadas segundo a correlação de Pearson e fator de inflação da variância (VIF). Em ocorrência de valores estimados de VIF igual a 1, foi considerado que não existia correlação entre os descritores; para os valores entre 1-5 a correlação foi considerada aceitável; e se o valor fosse maior que 10 considerou-se a existência de multicolinearidade e o modelo seria rejeitado [110]. As **Figura 8** e **9** mostram a matriz de correlação e nas **Tabelas 1** e **2** são exibidos o VIF.

**Figura 8:** Matriz de correlação entre os descritores selecionados pelo AG para o SaDB.



Fonte: próprio autor

**Figura 9:** Matriz de correlação entre os descritores selecionados pelo AG para o SathfrDB ((a) SathfrDB-2D e (b) SathfrDB-3D).



Fonte: próprio autor.

**Tabela 1:** Resultado do fator de inflação para a seleção SaDB.

| Seleção SaDB |       |          |        |           |              |          |
|--------------|-------|----------|--------|-----------|--------------|----------|
| Desc.        | nR07  | SpMax_X  | GATS5s | nArCOOH   | nRNR2        | nN(CO)2  |
| VIF          | 1,26  | 2,48     | 1,42   | 1,94      | 1,62         | 1,90     |
| Desc.        | nOHp  | nOHt     | C-018  | minsCH3   | B09[O-O]     | B06[O-O] |
| VIF          | 1,28  | 2,04     | 1,87   | 1,21      | 1,28         | 1,26     |
| Desc.        | qpmax | F04[N-O] | ALOGP2 | s3_phSize | CATS2D_03_DP |          |
| VIF          | 1,87  | 2,04     | 1,21   | 1,26      | 1,12         |          |

**Tabela 2:** Resultado do fator de inflação dos descritores para o SathfrDB

| Seleção SadhfrDB-2D |         |                  |                  |                  |                   |
|---------------------|---------|------------------|------------------|------------------|-------------------|
| Desc.               | SpMax_X | SpMax3_Bh(p)     | SpMin1_Bh(m)     | CATS2D_04_DL     |                   |
| VIF                 | 1,18    | 2,93             | 1,97             | 1,7              |                   |
| Seleção SadhfrDB-3D |         |                  |                  |                  |                   |
| Desc.               | SHP2    | CATS3D_05_<br>DL | CATS3D_03<br>_LL | CATS3D_08_<br>LL | WHALES100<br>_Rem |
| VIF                 | 2,12    | 1,47             | 2,56             | 2,62             | 1,76              |

Observa-se que os conjuntos de descritores selecionados não apresentam correlação de Pearson acima de 0,6, que indica, portanto, que não estão autocorrelacionados entre si. Além disso, os valores calculados de VIF correspondente são todos menores que 5, o que corrobora que não há multicolinearidade e os descritores são independentes uns dos outros. O problema de multicolinearidade entre os descritores atrapalha a seleção de variáveis que permitem a criação de modelos estatisticamente significativos e robustos [111]. Assim, a seleção feita por AG foi eficaz para a seleção das variáveis mais significativas, uma vez que a partir da seleção foi possível construir modelos preditivos, como é observado nas **Tabelas 3 e 4**.

### 5.1.2 Construção dos modelos

Modelos lineares e não lineares foram então desenvolvidos no presente estudo. Com os métodos de aprendizado de máquina LR e PR foi possível analisar se existe relação linear entre os descritores e a atividade biológica; e usando os métodos, SVM e MLP foi verificada a existência de relação não linear. Os resultados estatísticos dos modelos gerados para a matriz de descritores SaDB e para as matrizes SadhfDBr-2D e SadhfrDB-3D podem ser observados nas **Tabelas 3 e 4**, respectivamente.

**Tabela 3:** Parâmetros estatísticos obtidos a partir dos modelos desenvolvidos para o conjunto SaDB.

| Parâmetros estatísticos dos modelos - SaDB |       |      |      |       |      |      |       |      |      |
|--------------------------------------------|-------|------|------|-------|------|------|-------|------|------|
|                                            | $R^2$ | MAE  | RMSE | $Q^2$ | MAE  | RMSE | $P^2$ | MAE  | RMSE |
| <b>SVM</b>                                 | 0,68  | 0,36 | 0,47 | 0,61  | 0,42 | 0,58 | 0,65  | 0,41 | 0,51 |
| <b>MLP</b>                                 | 0,73  | 0,34 | 0,43 | 0,62  | 0,42 | 0,57 | 0,73  | 0,37 | 0,45 |
| <b>LR</b>                                  | 0,61  | 0,39 | 0,52 | 0,47  | 0,44 | 0,6  | 0,61  | 4,36 | 0,53 |
| <b>PR</b>                                  | 0,61  | 0,39 | 0,52 | 0,47  | 0,44 | 0,6  | 0,61  | 436  | 0,53 |

**Tabela 4:** Parâmetros estatísticos obtidos a partir dos modelos desenvolvidos para o conjunto SadhfrDB

| Parâmetros estatísticos dos modelos - SadhfrDB-2D |       |      |      |       |      |      |       |      |      |
|---------------------------------------------------|-------|------|------|-------|------|------|-------|------|------|
|                                                   | $R^2$ | MAE  | RMSE | $Q^2$ | MAE  | RMSE | $P^2$ | MAE  | RMSE |
| <b>SVM</b>                                        | 0,77  | 0,41 | 0,53 | 0,6   | 0,54 | 0,52 | 0,64  | 0,49 | 0,56 |
| <b>MLP</b>                                        | 0,72  | 0,47 | 0,56 | 0,61  | 0,58 | 0,59 | 0,64  | 0,44 | 0,58 |
| <b>PR</b>                                         | 0,73  | 0,44 | 0,57 | 0,61  | 0,56 | 0,58 | 0,65  | 0,34 | 0,59 |
| <b>LR</b>                                         | 0,73  | 0,44 | 0,57 | 0,61  | 0,56 | 0,58 | 0,65  | 0,34 | 0,59 |
| Parâmetros estatísticos dos modelos - SadhfrDB-3D |       |      |      |       |      |      |       |      |      |
|                                                   | $R^2$ | MAE  | RMSE | $Q^2$ | MAE  | RMSE | $P^2$ | MAE  | RMSE |
| <b>SVM</b>                                        | 0,77  | 0,36 | 0,53 | 0,56  | 0,54 | 0,67 | 0,76  | 0,37 | 0,5  |
| <b>MLP</b>                                        | 0,69  | 0,51 | 0,61 | 0,55  | 0,59 | 0,74 | 0,67  | 0,42 | 0,58 |
| <b>PR</b>                                         | 0,62  | 0,51 | 0,54 | 0,5   | 0,57 | 0,63 | 0,6   | 0,47 | 0,54 |

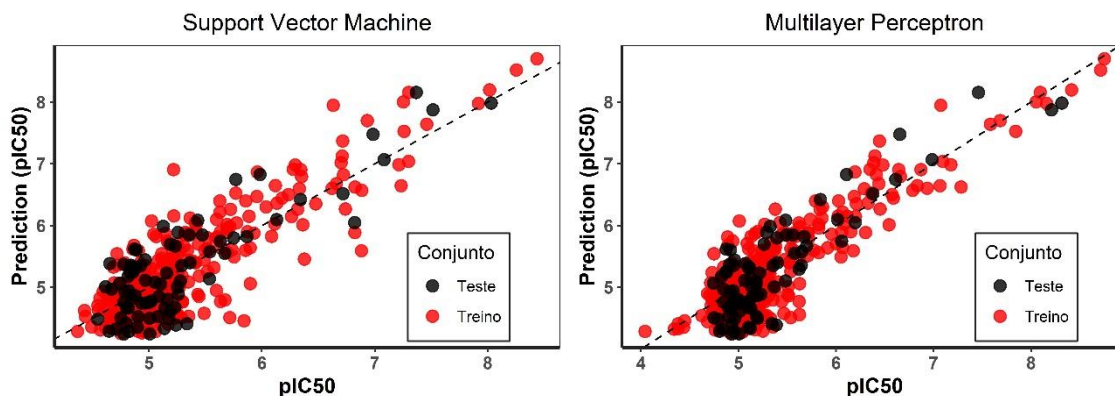
|    |      |      |      |     |      |      |     |      |      |
|----|------|------|------|-----|------|------|-----|------|------|
| LR | 0,62 | 0,51 | 0,54 | 0,5 | 0,57 | 0,63 | 0,6 | 0,47 | 0,54 |
|----|------|------|------|-----|------|------|-----|------|------|

$R^2$ = coeficiente de correlação.  $Q^2$ = coeficiente de correlação da validação cruzada.  $P^2$ = coeficiente de correlação da validação externa. MAE= mean absolute error (erro absoluto médio). RMSE= root mean squared error (erro quadrático médio). Fonte: próprio autor.

Resultados estatísticos são considerados satisfatórios para modelos que apresentam valores na validação cruzada  $R^2$  e  $Q^2$  superiores a 0,5 e 0,6, respectivamente [112]. Além disso, os valores entre as métricas  $R^2$  e  $Q^2$  devem ser menores que 0,3. Uma diferença maior que 0,3 entre essas métricas pode indicar presença de *outlier* no conjunto treino ou presença de descritores irrelevantes [113]. Desta forma, como se nota nas tabelas, a partir das métricas  $Q^2$  e  $R^2$ , os modelos gerados com os algoritmos LR e PR a partir da matriz SaDB não apresentam capacidade preditiva, pois apresentaram valor estatístico de validação externa ( $Q^2$ ) menor que o adequado [112]. Sendo assim, os modelos lineares para a matriz SaDB foram desconsiderados para as demais avaliações. Já os demais modelos obtidos podem ser considerados, pelas métricas  $R^2$  e  $Q^2$ , estáveis e robustos.

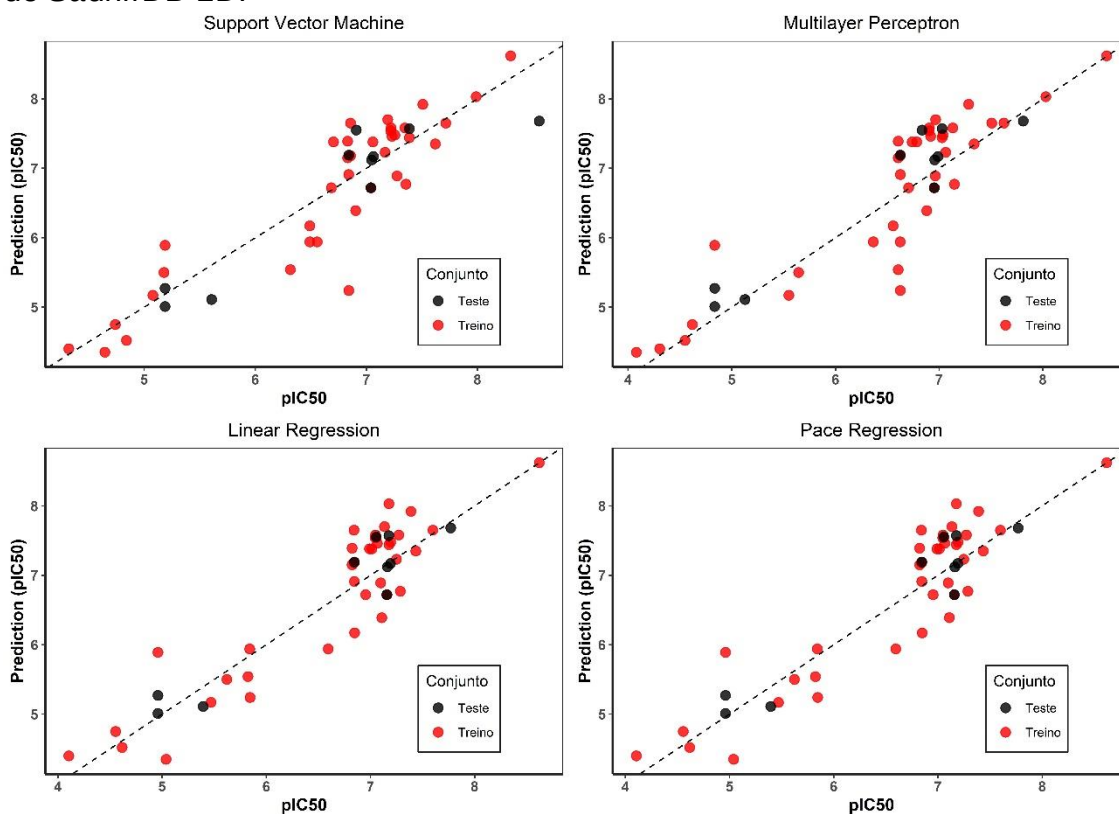
Do mesmo modo, um modelo QSAR, além de ser estável e robusto (avaliado por validação cruzada interna), também deve ser confiável em previsões de grupos de estruturas químicas que não fazem parte do grupo treino. Portanto, o coeficiente de correlação de um conjunto externo,  $P^2$ , foi calculado. Este deve apresentar valor maior ou igual a 0,6 [62,112]. Assim, observa-se que todos os valores de  $P^2$  obtidos nos modelos estão de acordo. Na **Figura 10-12** estão apresentados os gráficos de dispersão obtidos a partir dos valores reais de atividade vs os valores previstos para cada modelo.

**Figura 10:** Relação dos valores reais vs os valores previstos para os modelos do SaDB.

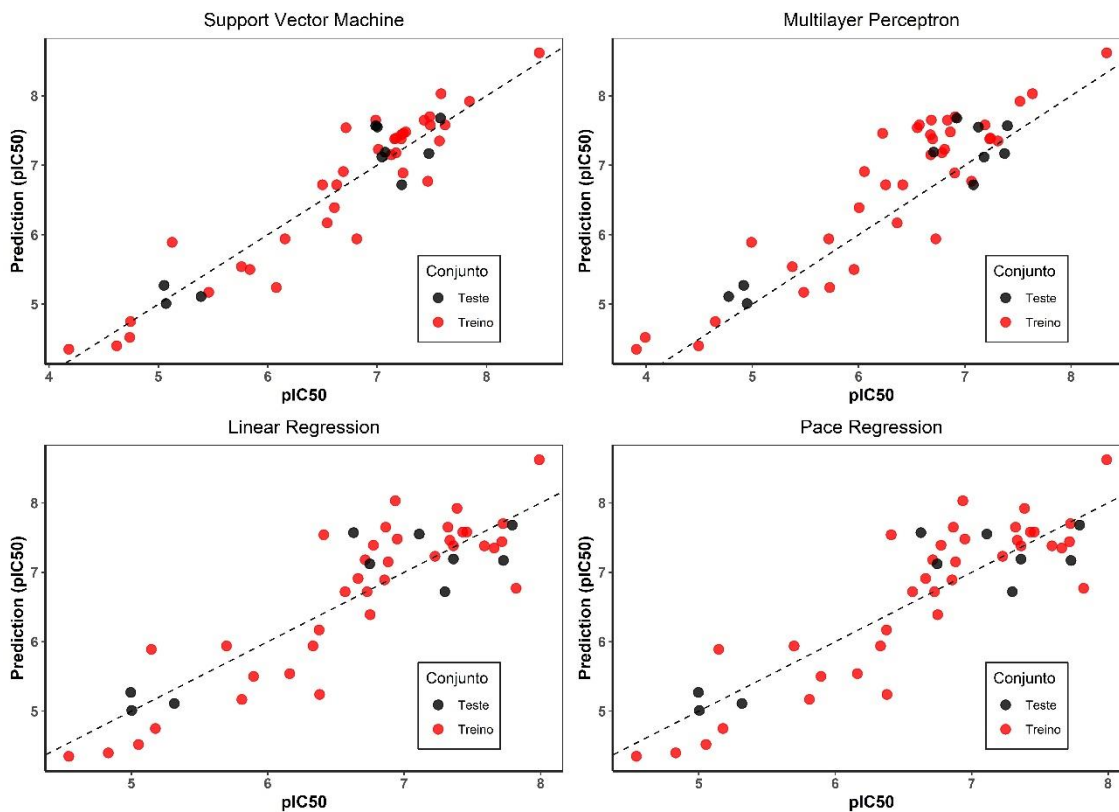


Fonte: próprio autor

**Figura 11:** Relação dos valores reais vs os valores previstos para os modelos do SadhfrDB-2D.



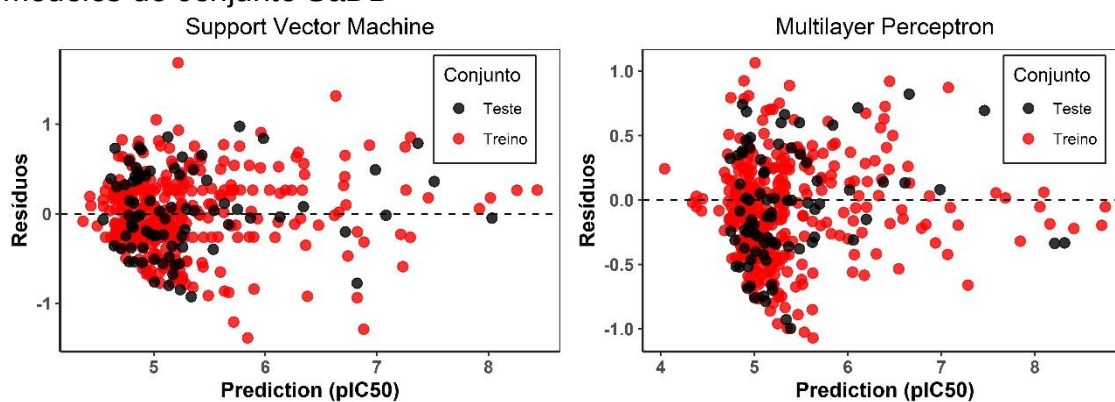
**Figura 12:** Relação dos valores reais vs os valores previstos para os modelos do SadhfrDB-3D.



Fonte: próprio autor

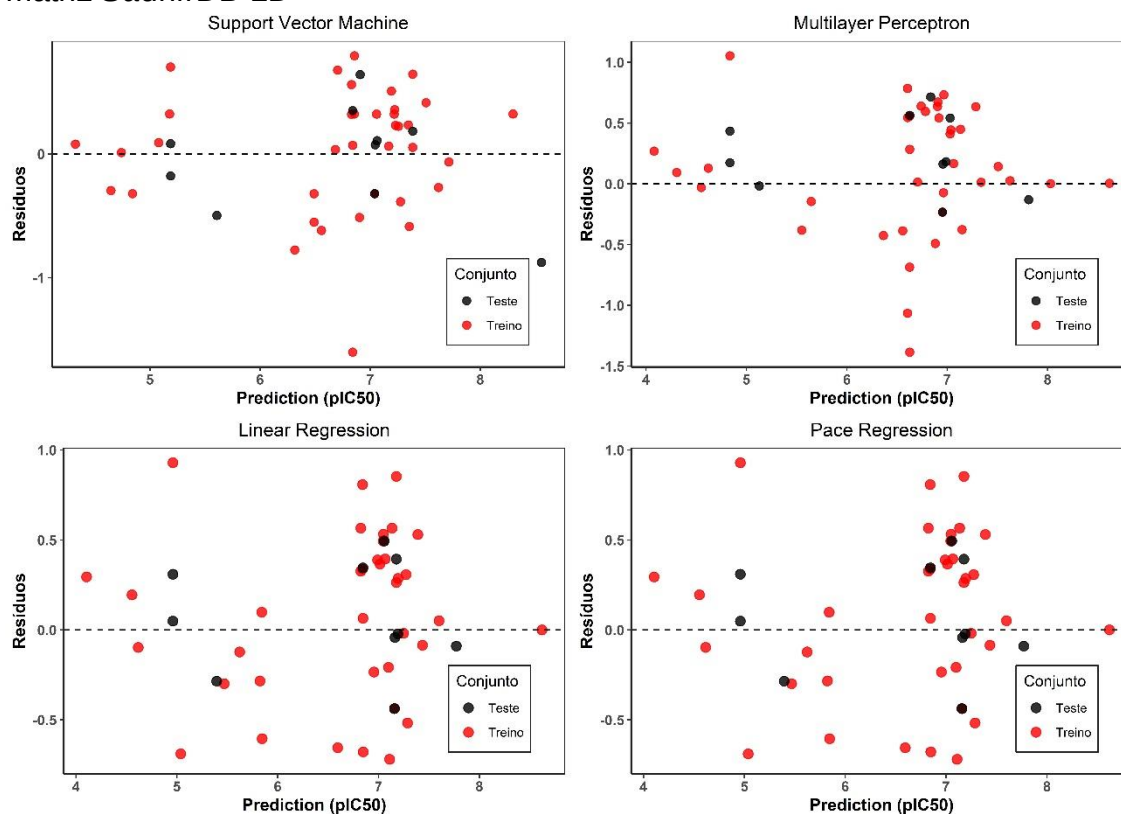
Nas **Figuras 13 - 15**, observa-se o gráfico de dispersão do  $pIC_{50}$  real em relação aos resíduos obtidos a partir das previsões. A distribuição aleatória dos resíduos em torno de zero indica que não há erros sistemáticos e de falta de ajuste dos modelos [114]. Entre os modelos desenvolvidos, a partir do conjunto SaDB observam-se menores erros associados para o modelo desenvolvido pelo algoritmo MLP. Para os modelos desenvolvidos a partir dos conjuntos SadhfrDB-2D e SadhfrDB-3D os menores resíduos associados são observados para os modelos lineares e SVM, respectivamente.

**Figura 13:** Relação dos valores previstos vs os resíduos dos modelos para os modelos do conjunto SaDB

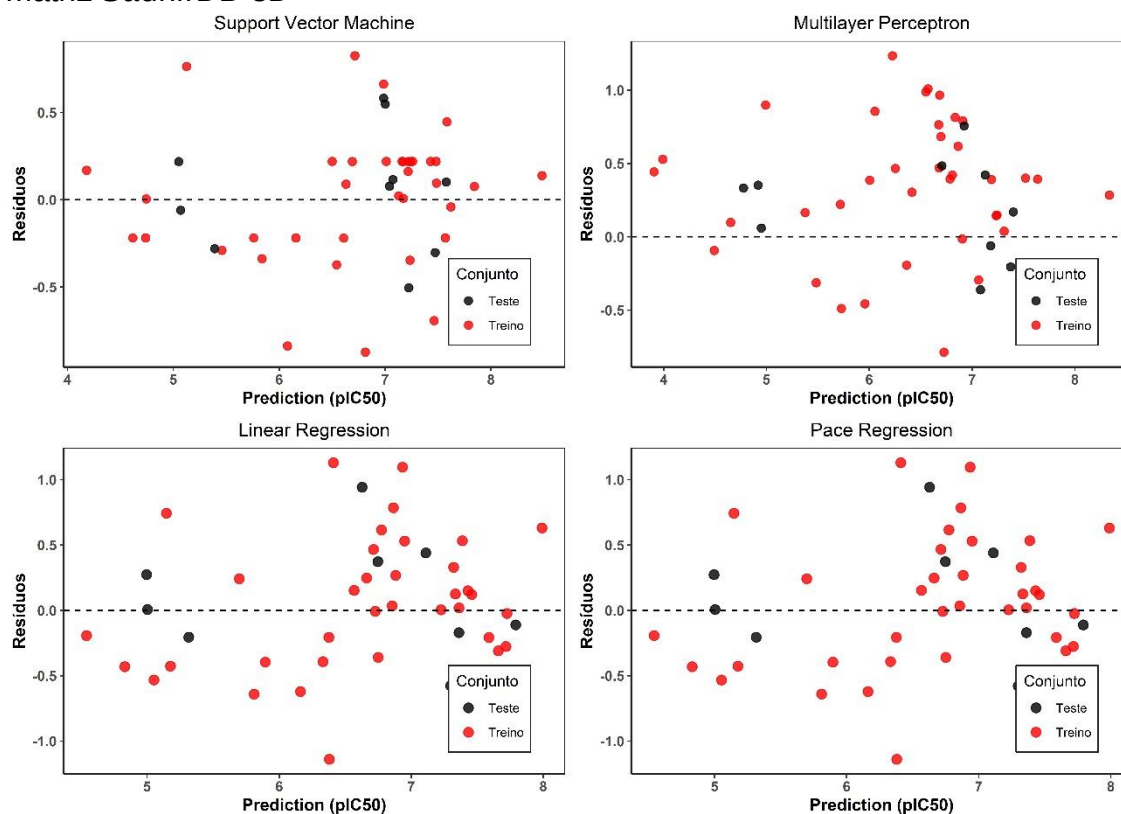


Fonte: próprio autor

**Figura 14:** Relação dos valores previstos vs os resíduos dos modelos para a matriz SadhfrDB-2D



**Figura 15:** Relação dos valores previstos vs os resíduos dos modelos para a matriz SadhfrDB-3D



Fonte: próprio autor

Os resultados obtidos nos modelos indicam resultados estatísticos satisfatórios em relação às métricas de avaliação interna e externa ( $R^2$ ,  $Q^2$  e  $P^2$ ). Adicionalmente, para atribuir aos modelos maior confiabilidade, a capacidade preditiva dos modelos foi avaliada utilizando Critérios de Aceitabilidade dos modelos que estabelecem critérios estatísticos para que um modelo QSAR seja considerado verdadeiramente preditivo. Os critérios de aceitabilidade desenvolvidos por Golbraikh e Tropsha (2002) é uma ferramenta que corrobora com os resultados estatísticos obtidos para os modelos [115]. Os resultados dos critérios de aceitabilidade para os modelos estão apresentados na **Tabela 5**

**Tabela 5:** Resultados estatísticos dos Critérios de Aceitabilidade para os modelos.

| <b>Resultados dos Critérios de Aceitabilidade- SaDB</b>         |       |       |                                      |                                       |                         |                          |
|-----------------------------------------------------------------|-------|-------|--------------------------------------|---------------------------------------|-------------------------|--------------------------|
|                                                                 | $R^2$ | $Q^2$ | $\frac{(R^2 - R_o^2)}{R^2} \leq 0,1$ | $\frac{(R^2 - R_o'^2)}{R^2} \leq 0,1$ | $0,85 \leq k \leq 1.15$ | $0,85 \leq k' \leq 1.15$ |
| SVM                                                             | 0,655 | 0,649 | 0,362                                | 0,003                                 | 1,011                   | 0,98                     |
| MLP                                                             | 0,737 | 0,728 | 0,242                                | 0,009                                 | 0,992                   | 1,001                    |
| <b>Resultados dos Critérios de Aceitabilidade- SadhfrDB -2D</b> |       |       |                                      |                                       |                         |                          |
| SVM                                                             | 0,673 | 0,644 | 0,122                                | 0,009                                 | 0,977                   | 1,016                    |
| MLP                                                             | 0,697 | 0,641 | 0,013                                | 0,081                                 | 1,0                     | 0,992                    |
| LR                                                              | 0,708 | 0,652 | 0,025                                | 0,046                                 | 0,977                   | 1,016                    |
| PR                                                              | 0,712 | 0,652 | 0,025                                | 0,046                                 | 0,977                   | 1,016                    |
| <b>Resultados dos Critérios de Aceitabilidade- SadhfrDB -3D</b> |       |       |                                      |                                       |                         |                          |
| SVM                                                             | 0,786 | 0,761 | 0,008                                | 0,027                                 | 0,99                    | 1,004                    |
| MLP                                                             | 0,782 | 0,677 | 0,002                                | 0,049                                 | 1,041                   | 0,955                    |
| LR                                                              | 0,707 | 0,606 | 0,001                                | 0,127                                 | 0,983                   | 1,008                    |
| PR                                                              | 0,707 | 0,606 | 0,001                                | 0,127                                 | 0,983                   | 1,008                    |

Fonte: próprio autor.

Os resultados obtidos mostram que os modelos possuem potencial preditivo e podem ser usados na descoberta de novas substâncias ativas. Os algoritmos MLP para o SaDB, LR para SadhfrDB-2D e SVM para SadhfrDB-3D se mostraram ligeiramente mais preditivo que os demais. Contudo, para confirmar que os modelos não foram gerados por correlação aleatória foi

realizado o teste de aleatorização de Y. Espera-se que os valores de  $R^2$  e  $Q^2$  obtidos no teste sejam menores que os valores obtidos nos modelos [116]. Os resultados para os melhores modelos de cada conjunto são apresentados na **Tabela 6**.

**Tabela 6:** Resultados obtidos no teste de randomização.

| Randomização de Y |        |             |             |             |             |             |             |
|-------------------|--------|-------------|-------------|-------------|-------------|-------------|-------------|
| Seleção           | Modelo | $R^2$       | MAE         | RMSE        | $Q^2$       | MAE         | RMSE        |
| SaDB              | MLP    | 0,151       | 0,579       | 0,676       | -0,169      | 0,497       | 0,667       |
|                   |        | $\pm 0,089$ | $\pm 0,053$ | $\pm 0,039$ | $\pm 0,087$ | $\pm 0,017$ | $\pm 0,025$ |
| SadhfrDB-2D       | LR     | 0,086       | 0,559       | 0,604       | -0,278      | 0,523       | 0,647       |
|                   |        | $\pm 0,112$ | $\pm 0,072$ | $\pm 0,065$ | $\pm 0,148$ | $\pm 0,059$ | $\pm 0,07$  |
| SadhfrDB-3D       | SVM    | 0,378       | 0,56        | 0,768       | -0,469      | 0,476       | 0,633       |
|                   |        | $\pm 0,106$ | $\pm 0,081$ | $\pm 0,076$ | $\pm 0,279$ | $\pm 0,111$ | $\pm 0,13$  |

\* $\pm$  Desvio Padrão

Fonte: próprio autor.

Como pode ser visto, os valores médios de  $R^2$  e  $Q^2$  obtidos para os modelos aleatórios foram baixos, confirmando que os modelos gerados não foram obtidos ao acaso e são estatisticamente significativos [117]. O teste de aleatorização garante a robustez dos modelos desenvolvidos [118]. Assim, os modelos QSAR criados a partir de cada conjunto de descritores foram usados para triar alcaloides que possam ser ativos no *S. aureus*. Para os conjuntos de descritores foi realizado uma análise da contribuição dos descritores na previsão da atividade.

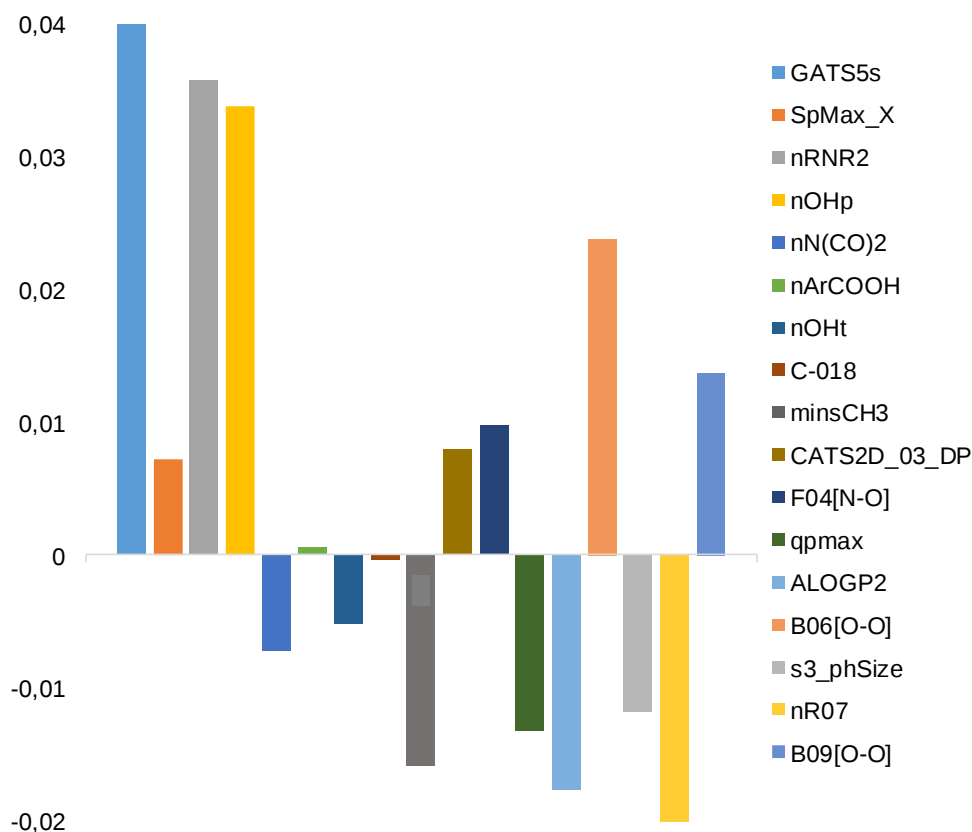
### 5.1.3 Análise dos descritores

#### 5.1.3.1 Seleção para o conjunto SaDB

Todos os 17 descritores aplicados para desenvolver os modelos trazem informações 2D das estruturas. Para entender a contribuição de cada descritor para a previsão da atividade do modelo, foram calculados os valores de SHAPE que quantificam a provável influência de uma determinada variável para a previsão do modelo [119]. Os valores podem apresentar sinais diferentes, valores de SHAPE com sinal positivo atribuídos ao descritor indicam que este contribui positivamente com a atividade e valores negativos significam que a

variável contribui pouco [120]. Na **Figura 16** é possível observar a contribuição média de cada descritor.

**Figura 16:** Contribuição média dos descritores selecionados para as previsões do modelo MLP.



Fonte: próprio autor.

O descritor com maior contribuição média, GATS5s, trata-se de um descritor de autocorrelação de Geary de atraso 5 ponderado pelo estado I. Esse descritor é da classe de descritores de Autocorrelação 2D e codifica informações de correlação espacial ponderada pelo estado eletropológico que correlaciona propriedades eletrônicas e estéricas da molécula [121]. Considera também a eletronegatividade de átomos próximos ou distantes [122]. O cálculo dos descritores GATS é dado pela equação 8.

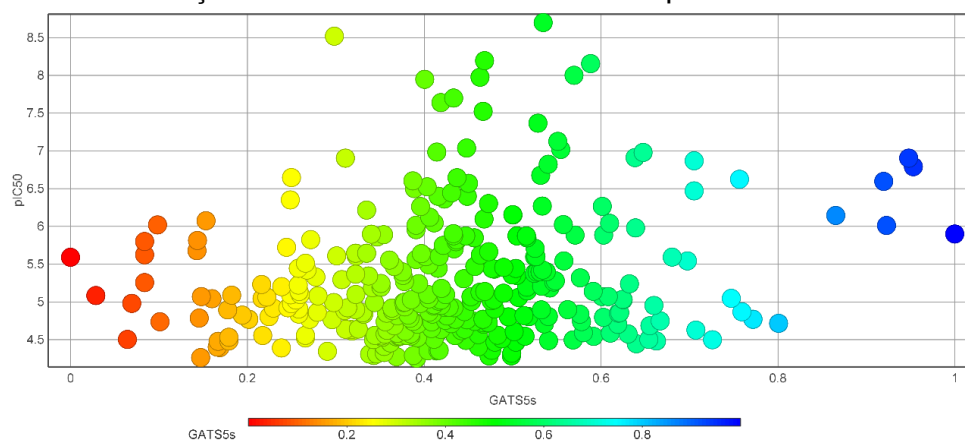
$$GATS(P_k, l) = \frac{N - 1}{4L} \frac{\sum_{ij} \delta_{ij} (p_{ki} - \bar{p}_k)(p_{kj} - \bar{p}_k)}{\sum_i (p_{ki} - \bar{p}_k)} \quad (8)$$

Onde  $l$  é a defasagem espacial;  $p_{ki}$  e  $p_{kj}$  são os valores da propriedade  $k$  do átomo  $i$  e  $j$ , respectivamente;  $P_k$  é o valor médio da propriedade  $k$ ;  $L$  é o

número de valores diferentes de zero na soma;  $N$  é o número de átomos na molécula; e  $\delta_{ij}(l, d_{ij})$  é uma função Dirac-delta [123,124].

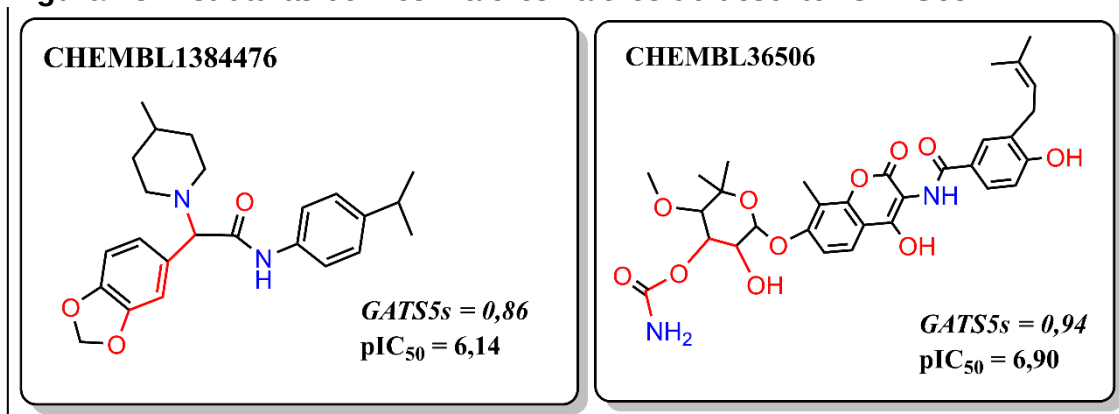
Na **Figura 17**, esse descritor foi correlacionado com o valor de  $\text{pIC}_{50}$ . Observa-se que as moléculas com maiores valores de  $\text{pIC}_{50}$  apresentam valores do descritor na faixa 0,3 a 0,6, aproximadamente. Descritores com maiores valores apresentam aproximadamente potenciais de atividade na faixa de 6 a 7.

**Figura 17:** Correlação entre o descritor GATS5s vs  $\text{pIC}_{50}$ .

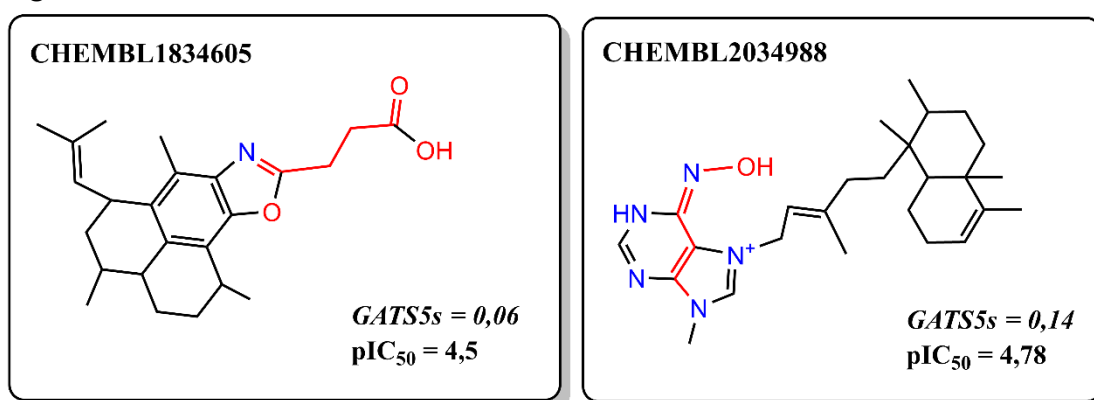


Fonte: próprio autor.

A relação entre o potencial de atividade e as possíveis características que influenciam no valor do descritor podem ser visualizadas na **Figura 18**, a qual apresenta as estruturas com maiores valores do descritor. Por outro lado, a **Figura 19** apresenta as estruturas com os valores mais baixos do descritor (em vermelho está um possível caminho que contribui para o aumento do descritor). A relação observada entre o  $\text{pIC}_{50}$  e as características das moléculas nas figuras sugere que os efeitos da eletronegatividade entre os átomos nas estruturas da **Figura 18** são mais relevantes que os efeitos observados nas moléculas da **Figura 19** para o aumento da atividade. Isso pode ser verificado pois as moléculas que apresentam maiores  $\text{pIC}_{50}$  possuem um número maior de átomos mais eletronegativos.

**Figura 18:** Estruturas com os maiores valores do descritor GATS5s.

Fonte: próprio autor

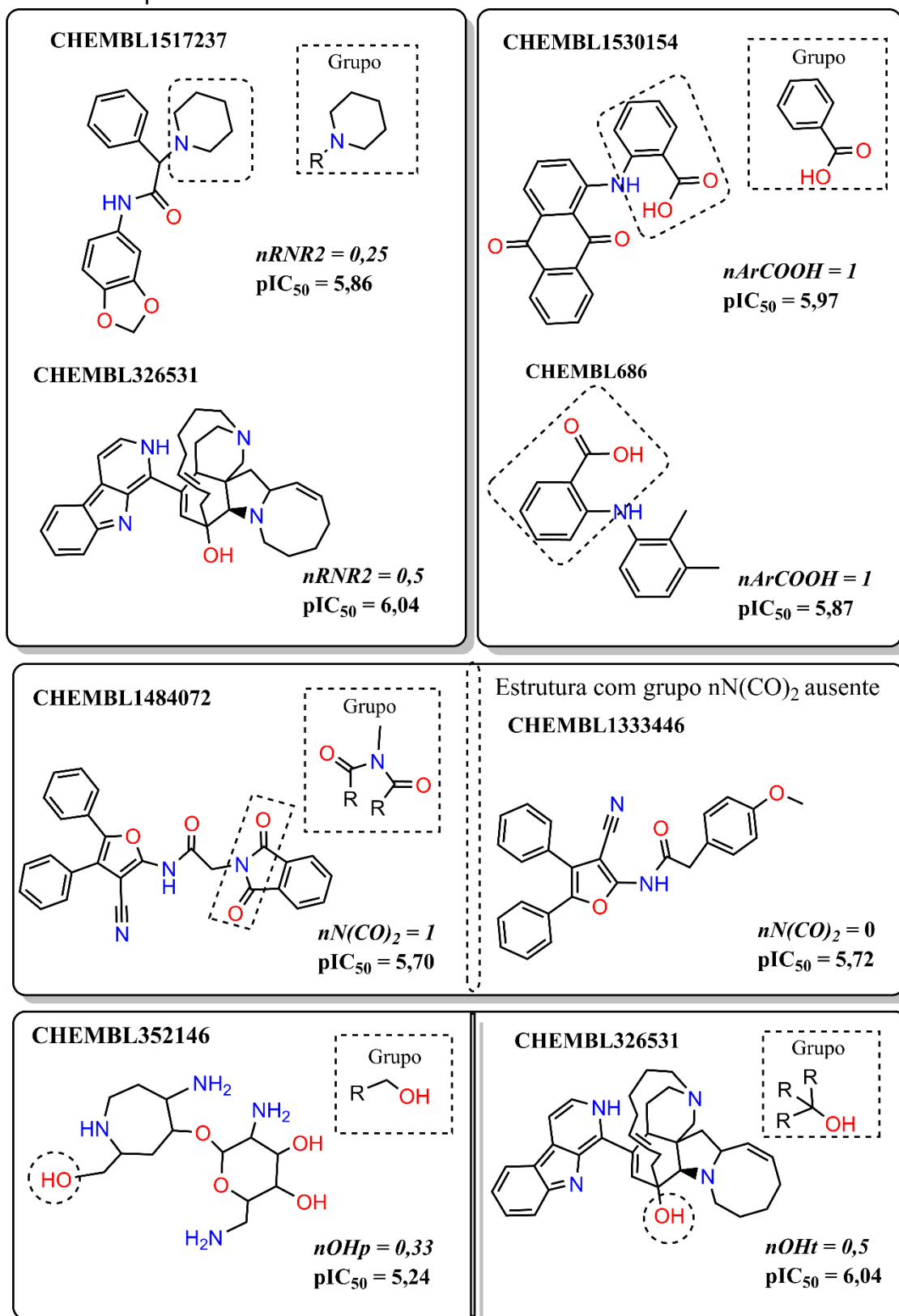
**Figura 19:** Estruturas com os menores valor do descritor GATS5s.

Fonte: próprio autor

O segundo descritor de maior contribuição, nRNR2, representa o número de aminas terciárias (alifáticas) e faz parte da classe de descritores de contagem de grupos funcionais. Outros descritores selecionados, como nN(CO)<sub>2</sub> (nº de imidas), nArCOOH (nº de ácidos carboxílicos aromáticos), nOHp (nº de álcool primário), nOHt (nº de álcool terciário) também fazem parte da classe de descritores de contagem de grupos. Na **Figura 20** observam-se algumas estruturas que apresentam características calculadas para os descritores.

Os descritores nN(CO)<sub>2</sub> e nArCOOH apresentaram menor frequência no conjunto treino. O potencial de atividade para as estruturas que apresentam grupos nN(CO)<sub>2</sub> em relação às que não apresentam, varia pouco, o que pode estar relacionado com sua contribuição negativa. Por outro lado, o descritor nArCOOH, ainda com sua baixa ocorrência apresentou contribuição para a atividade, visto que, pode apresentar relação com interações de hidrogênio que fazem parte das interações visualizadas entre ligante e receptor [125].

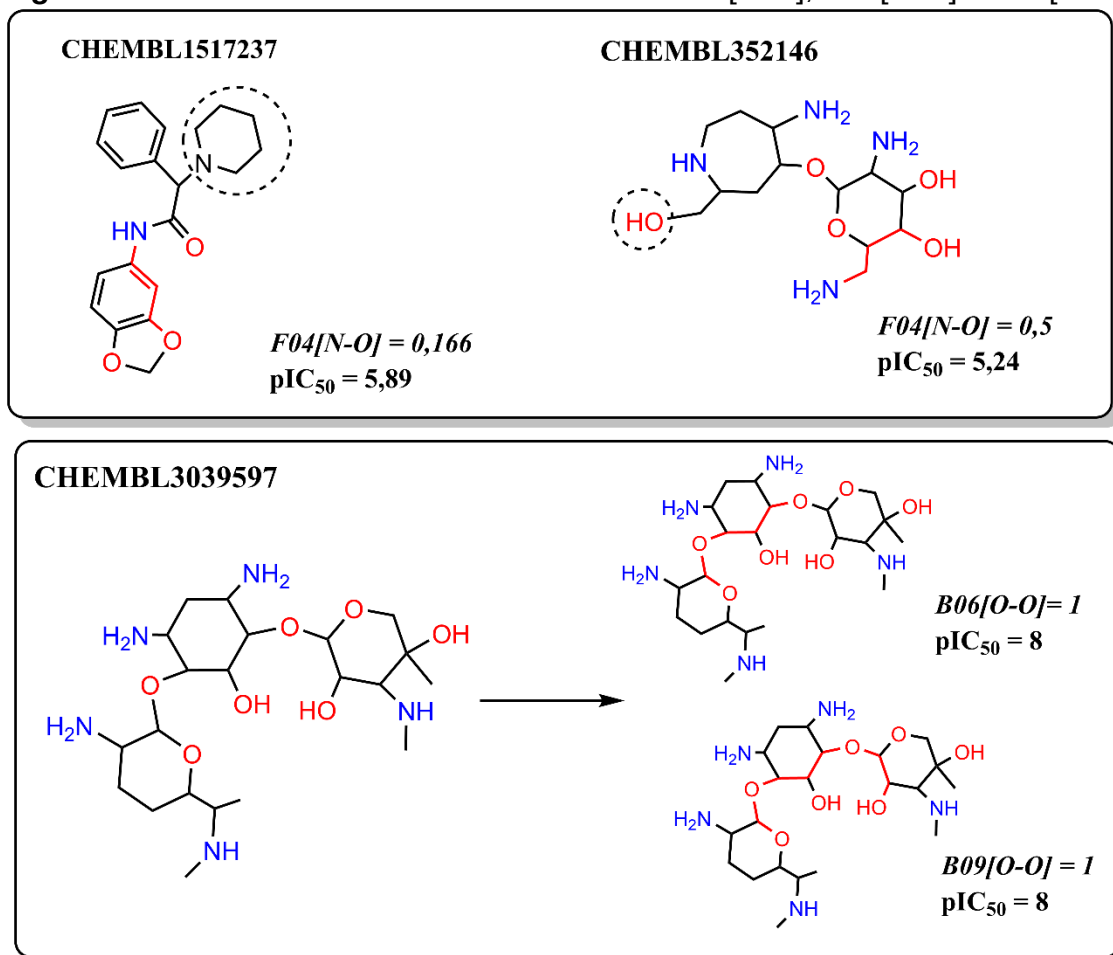
**Figura 20:** Estruturas com valores do descritor  $nRNR_2$ ,  $nN(CO)_2$ ,  $nArCOOH$ ,  $nOH_t$  e  $nOH_p$ .



Fonte: próprio autor

Os descritores F04[N-O], B06[O-O] e B09[O-O], que apresentam contribuições positivas, fazem parte da classe de descritores de contagem de pares de átomos. O descritor F04[N-O] calcula a frequência dos pares N-O na distância topológica 4. Os descritores B06[O-O] e B09[O-O] referem-se à presença/ausência de pares O-O na distância topológica 6 e 9, respectivamente. Na **Figura 21** estão apresentadas as estruturas que possuem as características codificadas pelos descritores, em que as ligações marcadas em vermelho indicam o possível caminho até os pares de átomos. As estruturas também carregam características codificadas por outros descritores, como nRNR2 e nOHp.

**Figura 21:** Estruturas com valores do descritor F04[N-O], B06[O-O] e B09[O-O].

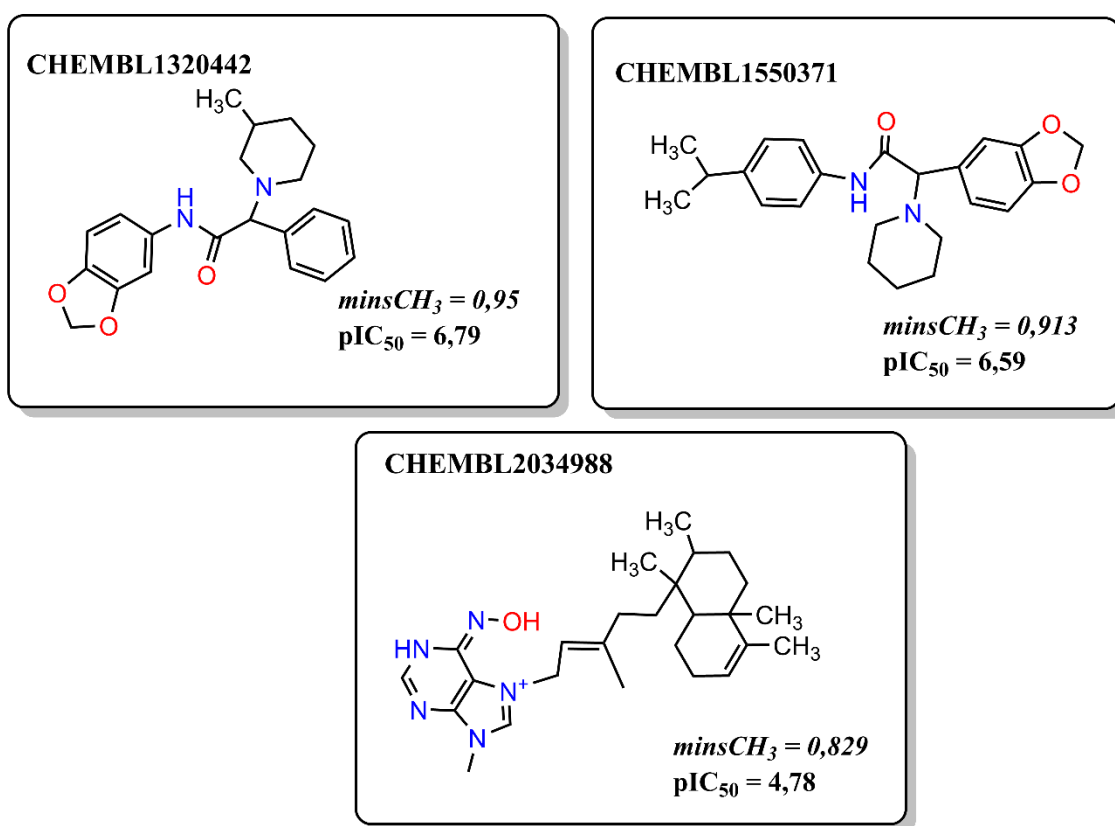


Fonte: próprio autor

O descritor minsCH3 indica O estado mínimo do CH<sub>3</sub>. Os índices de estado eletrotológico (estado E) foram introduzidos por Hall e Kier [126] para descrição de moléculas no nível atômico. O estado eletrotológico do grupo

CH<sub>3</sub> parece ser importante para interações hidrofóbicas entre ligante e receptor [125]. Esse descritor combina informações não apenas eletrônicas e topológicas do átomo analisado, mas também considera informações de estado de valência, características atômicas, acessibilidade de elétrons, estado eletropológico de átomos de hidrogênio e conectividade que são influenciados por todas as características estruturais de uma molécula [127,128]. Na **Figura 22** são apresentadas estruturas que indicam altos valores desse descritor. Esse descritor revelou contribuição média negativa.

**Figura 22:** Estruturas com valores altos do descritor *minsCH<sub>3</sub>*.

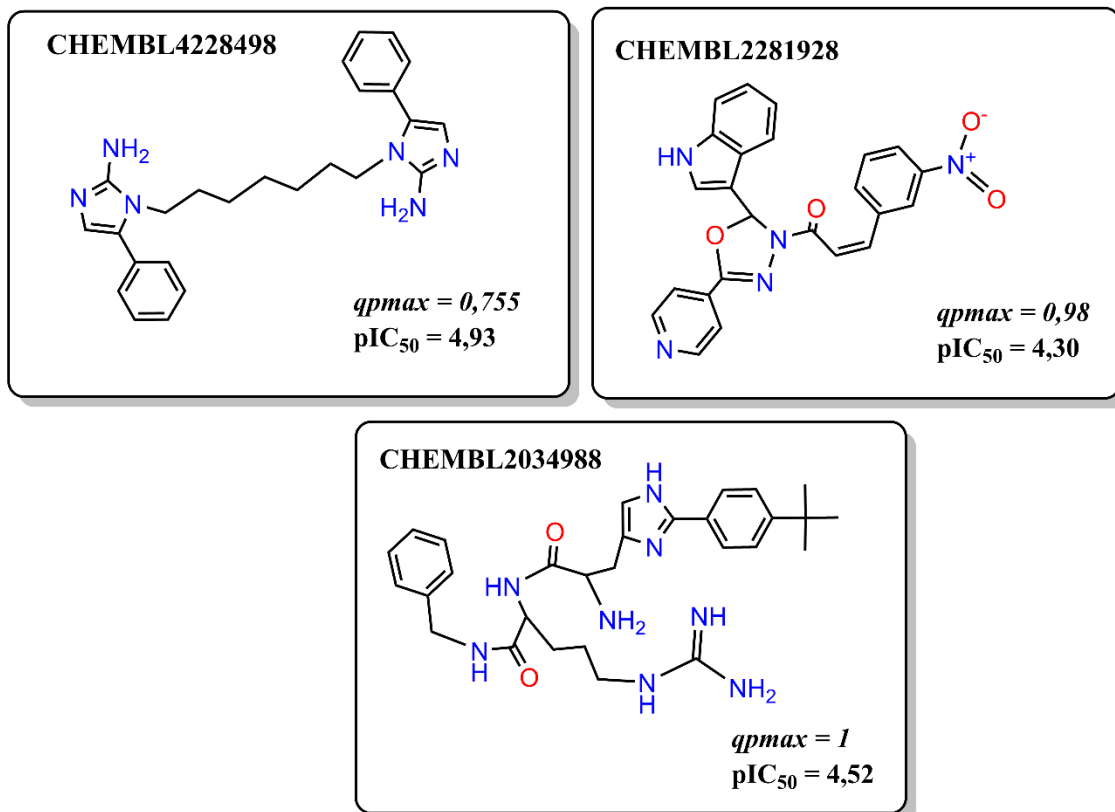


Fonte: próprio autor

O descritor nr07 é um descritor que traz informações do número de anéis de 7 membros de uma molécula. Esse descritor apresentou no conjunto treino pouca ocorrência. Para as estruturas que apresentavam tal característica, seus potenciais de atividade variaram de 4,37 a 6,64 M. O descritor qpmax traz informações da carga positiva máxima da molécula. É um descritor de carga atômica usado para descrever características eletrônicas de toda a molécula e de regiões particulares, como de átomos, ligações e fragmentos [129]. Na **Figura**

**23** pode ser observado as moléculas que apresentam valores altos do descritor, isso pode estar relacionado ao momento de dipolo das moléculas que influenciar a polaridade molecular em razão da separação de cargas induzidas. Ambos os descritores apresentaram contribuição negativa.

**Figura 23:** Estruturas que apresentam valores altos do descritor  $qp_{max}$ .

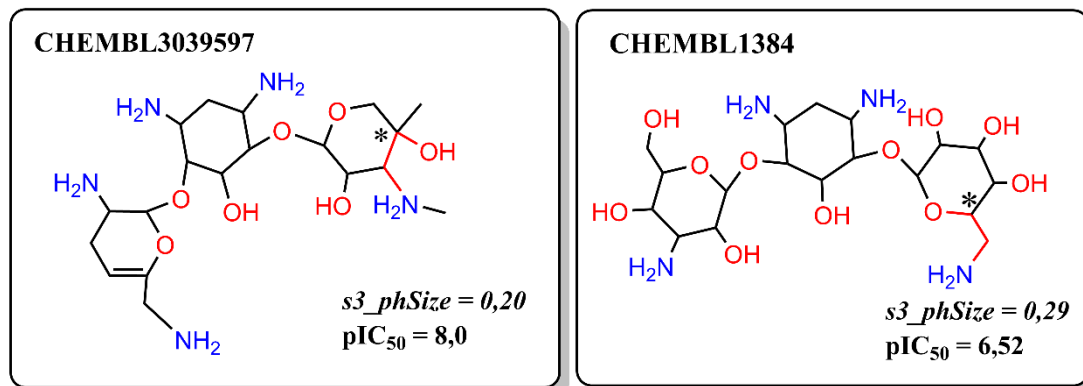


Fonte: próprio autor.

O descritor  $s3\_phSize$  é um descritor de quiralidade que traz informações do número de átomos de N/O compartilhados por um centro quiral normalizado a 3 ligações entre N/O. Em resumo, descritores de quiralidade são desenvolvidos com base em informações topológicas de grafos moleculares que incluem índices de conectividade molecular, índices de grupo de Zagreb e índices de carga topológica [115,130]. O cálculo do descritor se baseia nas seguintes etapas: (i) são calculados todos os caminhos topológicos do centro quiral até todos os átomos; (ii) são calculados os caminhos do centro quiral até os substituintes. Na **Figura 24** são mostradas duas estruturas que apresentam centros quirais normalizados a 3 ligações por nitrogênio entre os substituintes. A ligação marcada em vermelho indica um possível caminho aplicado no cálculo do descritor. Descritores de quiralidade são utilizados em estudos QSAR para

capturar propriedades locais do centro quiral que permite identificar a influência desses para a atividade biológica [131].

**Figura 24:** Estruturas com valores do descritor  $s3\_phSize$ .



Fonte: próprio autor.

Outros descritores selecionados, como CATS2D\_03\_DP da classe de descritores de pares farmacofórico 2D codificam informações de pares de átomos doadores positivos em uma distância de 3 ligações, C-18 da classe de descritores de fragmentos centrados nos átomos, onde C-18 é o fragmento  $=CHX$  ( $X = O, N, S, P, Se$  e halogênios) e ALOGP2 (coeficiente de partição octanol-água (AlogP) de Ghose-Crippen-Viswanadhan ao quadrado), que está relacionado a informações farmacocinéticas da molécula [132], apresentaram contribuições negativas.

Em resumo, os descritores selecionados indicaram a importância de informações estéricas e eletrônicas das moléculas, das distâncias topológicas nos pares [O-O] e [N-O], contribuições dos fragmentos RNR2 e OH, bem como, da influência de átomos eletronegativos que afetam o estado eletrotológico das moléculas para a contabilização das atividades inibitórias de *S. aureus*.

#### 5.1.3.2 Seleção para o conjunto SdhfrDB

Para interpretação da contribuição dos descritores serão usadas as equações fornecidas pelos modelos lineares. Nas equações 7 e 8 são apresentadas as equações fornecidas aplicando o algoritmo de regressão linear (LR) nas matrizes SdhfrDB-2D e SdhfrDB-3D, respectivamente.

$$pIC_{50} = (-1.3728 * SpMax\_X) + (3.2733 * SpMax3\_Bhp) \\ + (-2.2149 * SpMin1\_Bhm) + (1.3416 * CATS2D\_04\_DL) + 6.1455 \quad (7)$$

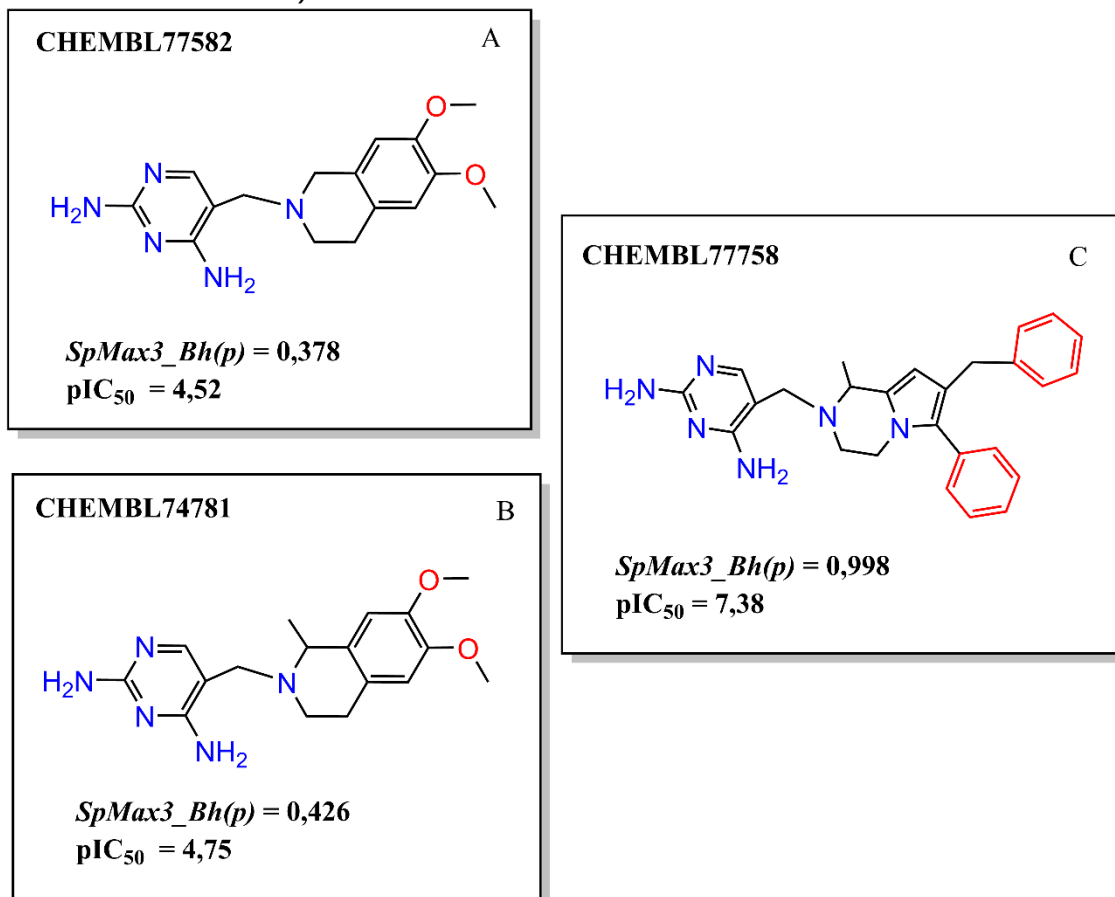
$$pIC_{50} = (-2.7043 * SHP2) + (1.2439 * CATS3D\_05\_DL) \\ + (2.3892 * CATS3D\_03\_LL) + (1.9108 * CATS3D\_08\_LL) \\ + (0.5984 * WHALES100\_Rem) + 6.5805 \quad (8)$$

#### 5.1.3.2.1 Subconjunto SadhfrDB-2D

$SpMax3\_Bh(p)$ ,  $SpMax\_X$  e  $SpMin1\_Bh(m)$  são descritores de autovalor relacionados a matriz de distâncias topológicas que capturam informações tanto da conectividade quanto atômicas de uma molécula. Em resumo, os descritores  $SpMax3\_Bh(p)$  e  $SpMin1\_Bh(m)$  são calculados por meio de diagramas moleculares contendo átomos de hidrogênio a partir da matriz de Burden,  $Bh(w)$ . A matriz  $Bh(w)$  é uma matriz de adjacência que leva em consideração a conectividade modificada por uma propriedade atômica relevante para a força de interação ligante-receptor, como massa, polarizabilidade e estado I [133]. Esses descritores apresentam relação com as ramificações de uma molécula [110,113] e podem solucionar problema de busca de similaridade estrutural [133].

O primeiro descritor na ordem de contribuição,  $SpMax3\_Bh(p)$ , chamado de autovalor principal ou raio espectral é um descritor de autovalor absoluto da matriz modificada de Burden –  $n^3$  ponderada pelas polarizabilidades relativas [110]. A polarizabilidade de uma molécula pode ser definida pela facilidade relativa de distorção da nuvem eletrônica quando exposta a um campo elétrico externo [134]. Esse descritor apresenta relação positiva com a atividade  $pIC_{50}$ . Assim, quanto maior o valor desse descritor maior será o  $pIC_{50}$ . Essa relação pode ser observada na **Figura 25** onde estão apresentados exemplos de estruturas com maior e menor valor desse descritor.

**Figura 25:** Estruturas que apresentam valores baixos e alto do descritor  $SpMax\_Bh(p)$ . (Estrutura A e B apresentam valores baixos do descritor e estrutura C valor alto).

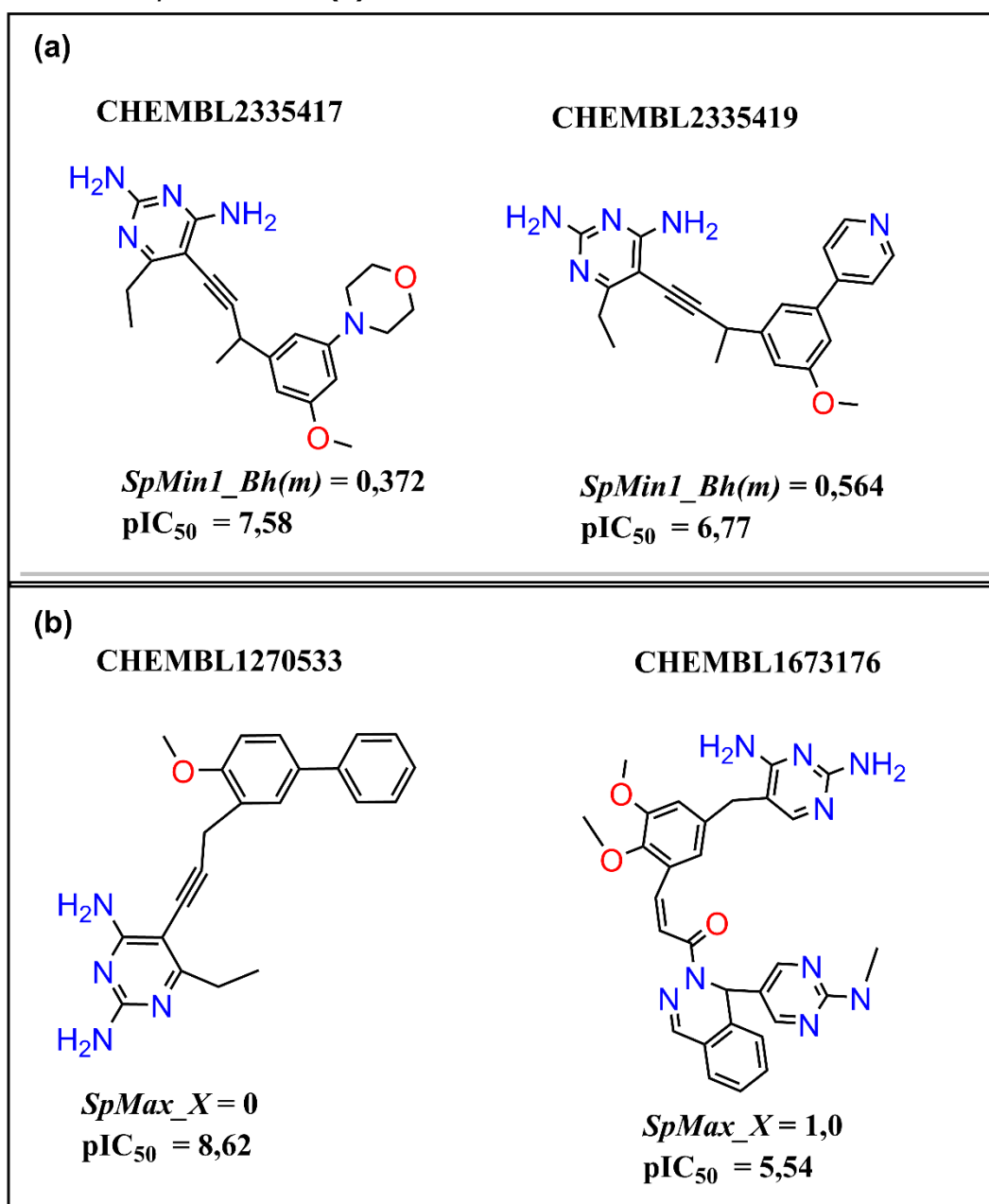


Fonte: próprio autor

A diferença entre as estruturas A e B é observada pela presença de uma ramificação ( $-CH_3$ ) a 3 ligações de distância do grupo 4 diaminopirimidina na estrutura B que resultou em um aumento na quantidade de elétrons e um pequeno aumento no valor do descritor. O maior valor para o descritor é observado para a estrutura C. Essa apresenta um aumento da nuvem eletrônica verificado por grupos aromáticos que aumentam a polarizabilidade da molécula. O grupo aromático fenila, forma um sistema de ligações  $\pi$  conjugadas que com a união de grupos doadores e retiradores de elétrons formam deformação da nuvem eletrônica e quanto maior essa deformação mais polarizável e maior a possibilidade de interação entre ligante-receptor [135], isso pois a polarizabilidade de uma molécula interfere nas forças intermoleculares, essas por sua vez, são fundamentais entre um fármaco e seu receptor [136].

Por outro lado, os descritores  $SpMin1\_Bh(m)$  e  $SpMax\_X$  apresentam contribuição negativa com a atividade. Dessa forma, o aumento na magnitude do descritor resultará em uma diminuição no valor de  $pIC_{50}$ . O descritor  $SpMin1\_Bh(m)$  é o menor autovalor absoluto da matriz modificada de Burdem – nº 1 ponderada pela massa e o descritor  $SpMax\_X$  é o autovalor principal da matriz chi e traz informações máximas de conexão. Na **Figura 26** pode ser vista a relação do descritor  $SpMin1\_Bh(m)$  e  $SpMax\_X$  com o  $pIC_{50}$ .

**Figura 26:** Relação do  $pIC_{50}$  com os descritores  $SpMin1\_Bh(m)$  em **(a)** e relação do descritor  $SpMax\_X$  em **(b)**.



Fonte: próprio autor.

A diferença nas estruturas pode ser observada pela introdução de átomos maiores nas moléculas, à medida que as ramificações ou átomos maiores são adicionados, as massas moleculares aumentam e a atividade diminui, o que sugere que o aumento da massa molecular tem uma relação negativa com atividade biológica da estrutura. O tamanho da molécula também influencia diretamente no valor do descritor *SpMax\_X*, pois moléculas maiores, com ramificações apresentam maior diversidade de conexões e maiores valores do descritor.

Os descritores *CATS* pertencem ao grupo de descritores de vetores de correlação que codificam informações sobre frequências de pares de átomos [137]. A classe de descritores *CATS2D* são altamente úteis e levam em consideração a distância topológica entre diferentes características farmacofóricas [138]. O cálculo geral dos descritores *CATS2D* é dado da seguinte forma: (i) obtenção do grafo molecular não ponderado e sem hidrogênio; (ii) os tipos de átomos são atribuídos ao grafo molecular; (iii) cálculo das distâncias; (iv) cálculo da representação do vetor de correlação *CVR*, dada pela equação 9:

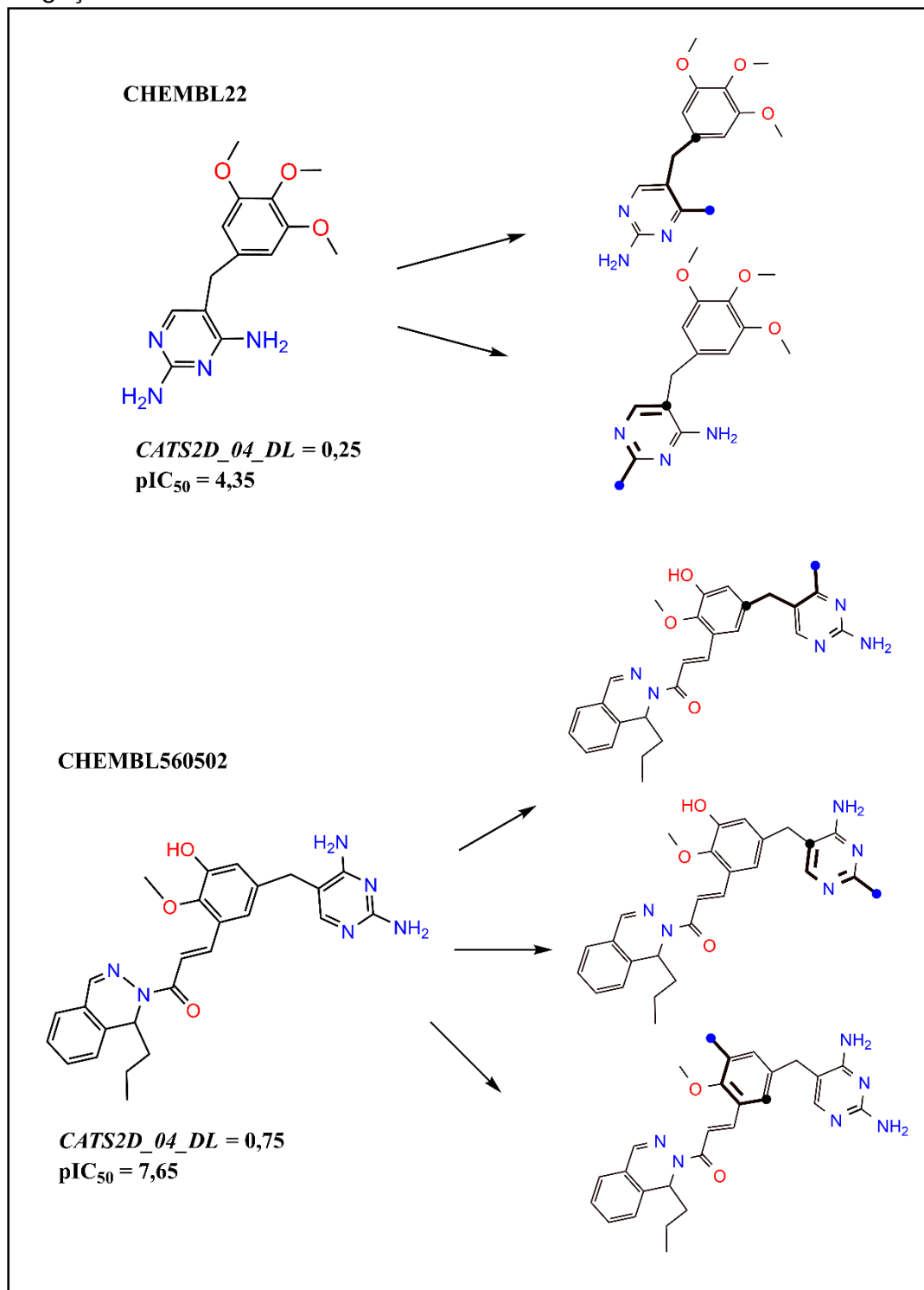
$$CVR_d^T = \frac{1}{A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij,d}^T \quad (9)$$

Onde, *T* é o par de tipos de átomos, *d* é o comprimento do caminho, *A* é o número de átomos diferentes de hidrogênio e  $\delta_{ij,d}^T$  é o delta de Kronecker que avalia como 1 se existe um par *T* e 0 caso contrário.

O descritor *CATS2D\_04\_DL* traz informações dos pares de átomos do tipo doador de ligações de hidrogênio e lipofílicos em distâncias topológicas de 4 ligações. Este apresenta contribuição positiva para a atividade biológica. Neste caso, as características da força doadora de ligações de hidrogênio e a lipofilicidade separadas na distância topológica quatro, aumentam o valor de *pIC*<sub>50</sub>. Na **Figura 27**, a relação desse descritor com a atividade é comparada. As

ligações destacadas indicam a distância entre os pares doador-lipofílico que contribuem no cálculo do descritor.

**Figura 27:** Relação do  $pIC_{50}$  com o descritor CATS2D\_04\_DL os pontos pretos indicam centros lipofílicos e azuis os centros doadores de hidrogênio que estão a 4 ligações de distância.



Fonte: próprio autor.

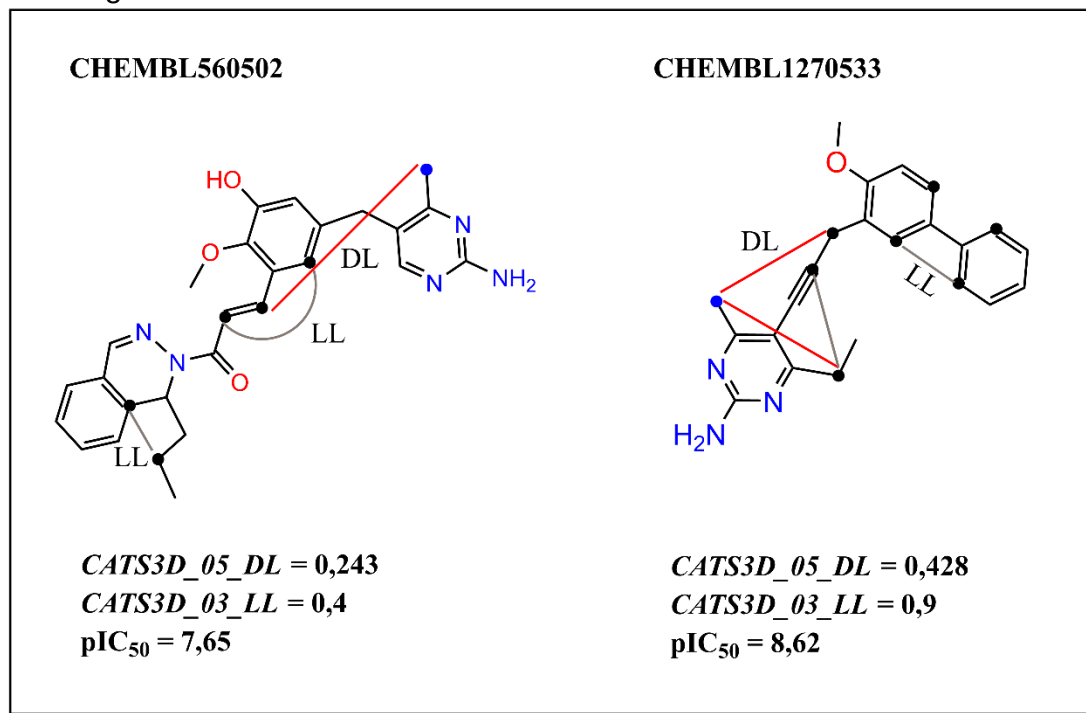
A lipofilicidade de uma molécula é uma propriedade muito importante relacionada ao transporte de fármacos no corpo, que incluem absorção intestinal e permeabilidade da membrana, formas de ligações a proteínas e distribuição entre diferentes tecidos. De maneira geral, sabe-se que as propriedades estruturais e físico-químicas codificadas pelos descritores *SpMax3\_Bh(p)* e CATS2D\_04\_DL têm um papel significativo que explicam sobre absorção intestinal de uma molécula que apresenta dependência da lipofilicidade com outros fatores, como tamanho da molécula, ligações de hidrogênio, hidrofilicidade e grau de ionização [107].

#### 5.1.3.2.2 Subconjunto SadhfrDB-3D

Os descritores moleculares CATS3D\_05\_DL, CATS3D\_03\_LL e CATS3D\_08\_LL pertencem ao grupo de descritores farmacóforos que usam distâncias euclidianas [137]. Os descritores CATS3D são uma extensão dos descritores CATS2D, porém em um espaço tridimensional. O cálculo do vetor de correlação é calculado análogo à equação 8, desconsiderando distâncias entre átomos idênticos (onde  $i=j$ ) [139]. CATS3D\_05\_DL vincula informações sobre pares de átomos doador-lipofílico a uma distância de 5 Å, enquanto os descritores CATS3D\_03\_LL e CATS3D\_08\_LL vinculam informações de pares de átomos lipofílico-lipofílico a 3 e 8 Å de distância, respectivamente.

A contribuição positiva apresentada pelos descritores CATS3D\_03\_LL, CATS3D\_08\_LL e CATS3D\_05\_DL indica que a atividade pode ser intensificada pelo aumento desses descritores, sugerindo que as características, lipofilicidade e força doadora de ligações de hidrogênio em inibidores da SaDHFR podem aumentar a atividade das moléculas estudadas. CATS3D\_03\_LL possui a maior contribuição para a atividade. Assim, pares de átomos relacionadas a uma distância de ligação menor apresentam maiores possibilidades de contribuições para o aumento da atividade. Na **Figura 28** estão ilustrados exemplos de pares doador-lipofílico (espaçado em vermelho) e pares lipofílico-lipofílico (espaçado em cinza) com seus valores dos descritores.

**Figura 28:** Relação do  $pIC_{50}$  com os descritores  $CATS3D\_05\_DL$  e  $CATS3D\_03\_LL$ . Os pontos pretos indicam centros lipofílicos e azuis doadores de hidrogênio.



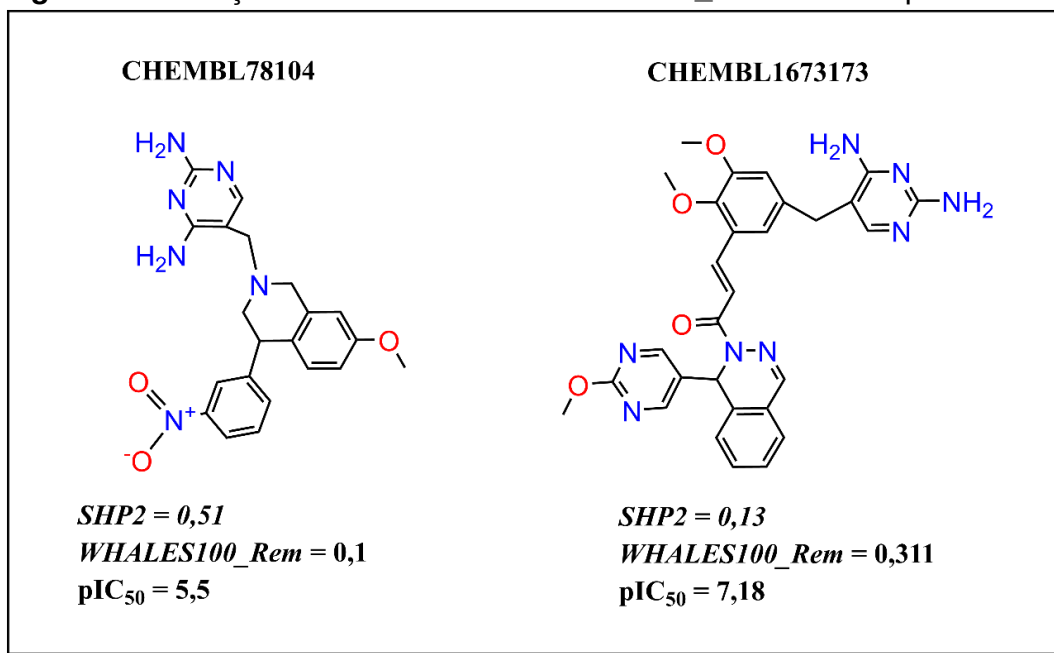
Fonte: próprio autor

O descritor SHP2 é um descritor da classe de perfis moleculares de Randic, que leva em consideração a média de perfil de vários descritores inferiores derivados da distância interatômica dos átomos na periferia de uma molécula [140]. Introduzido pelo químico e matemático Milan Randic [141], são baseados na teoria dos grafos e envolve no cálculo várias características moleculares como: forma molecular, tamanho e extensão. Tais características influenciam fenômenos de entropia e capacidade de interação entre ligante e o receptor [142]. O coeficiente de contribuição negativa desse descritor destaca sua relação negativa com o  $pIC_{50}$ , logo, quanto maior o valor desse descritor menor será o  $pIC_{50}$ . Essa relação pode ser observada na **Figura 29**.

Os descritores moleculares *WHALES* (*Weighttted Holistic Atom Localization and Entity Shape*) capturam informações de distâncias geométricas de interação entre os átomos, formato molecular e distribuição de carga parcial [143]. O descritor *WHALES100\_Rem*, codifica informações de frequência de afastamento, calculado em percentil a partir de distâncias interatômicas calculadas para dois pares de átomos diferentes de hidrogênio. O descritor

*WHALES100\_Rem* é computado de uma matriz de covariância centrada no átomo utilizando as coordenadas, cargas parciais e densidade dos átomos. Desta matriz de covariância, a distância de Mahalanobis normalizada é calculado e por fim é obtido o índice de afastamento. Esse descritor apresenta uma contribuição positiva para o aumento da atividade [144]. Na **Figura 29**, é observado que estruturas moleculares maiores apresentam valores do descritor *WHALES100\_Rem* maiores, pois moléculas maiores apresentam distâncias interatômicas maiores.

**Figura 29:** Relação dos descritores *WHALES100\_Rem* e *SPH2* para a atividade.



Fonte: próprio autor.

Assim, as combinações dos descritores codificam informações estéricas e eletrônicas importantes das moléculas, relacionadas ao tamanho molecular, ramificação, polarizabilidades, polaridade e lipofilia, que podem interferir nas interações hidrofóbicas e polares entre o ligante e o receptor SaDHFR.

## 5.2 Triagem virtual baseada em ligantes

O banco de triagem com 1200 estruturas de alcaloides foi verificado quanto ao domínio de aplicabilidade estabelecido anteriormente pelos descritores selecionados para desenvolver os modelos. Isso deve ser feito, pois

a previsão de uma molécula que esteja fora do espaço químico estabelecido não é confiável e resultado de uma extrapolação [132,145].

Assim, inicialmente foi verificado o DA estabelecido pelos descritores usados para desenvolver os modelos do conjunto SaDB. A cobertura foi de 94,5% de um total de 1148 alcaloides, assim 1080 foi a quantidade de estruturas em que as atividades foram previstas. Destas, 836 moléculas, que apresentaram probabilidade de serem ativas maior ou igual a 50%,  $pIC_{50} \geq 5$ , foram selecionadas para passarem por uma nova triagem, no intuito de descobrir as estruturas ativas na SaDHFR.

Antes da triagem, essas estruturas passaram mais uma vez pelo DA, agora estabelecido pelos descritores usados para desenvolver os modelos dos subconjuntos SadhfrDB-2D e SadhfrDB-3D. Os resultados das estruturas que passaram no DA são apresentados na **Tabela 7**.

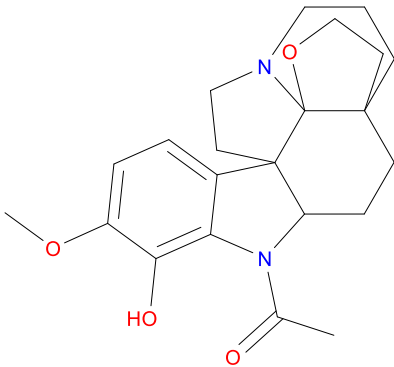
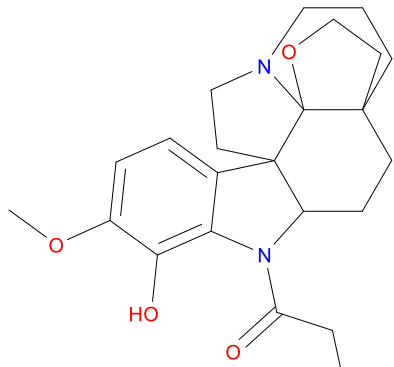
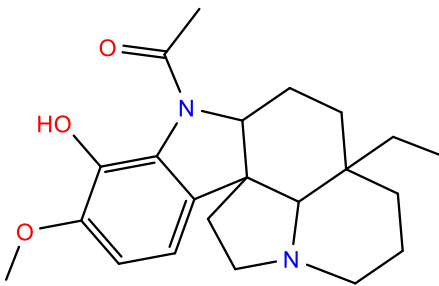
**Tabela 7:** Número de estruturas que fazem parte do DA

| N. de Estruturas | Subconjuntos | N. de estruturas que passaram pelo DA | Cobertura |
|------------------|--------------|---------------------------------------|-----------|
| 836              | SadhfrDB-2D  | 376                                   | 45,0%     |
|                  | SadhfrDB-3D  | 642                                   | 76,8%     |

Os valores das atividades dos alcaloides que passaram pelo filtro do DA foram previstos separadamente para cada conjunto de estruturas. Em seguida, para selecionar os alcaloides com a maior probabilidade de serem ativos contra SaDHFR dos dois modelos, uma análise de consenso foi realizada. Os modelos de consenso, são usados para aumentar a confiabilidade das previsões de modelos independentes [146–148].

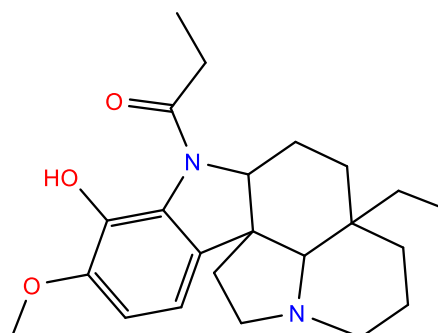
A análise de consenso resultou na seleção de 28 alcaloides com probabilidade de atividade contra SaDHFR. Esses alcaloides apresentam atividade nas faixas dos medicamentos iclaprim e TMP. Os 15 alcaloides com maiores  $pIC_{50}$  ativo, estão disponíveis na **Tabela 8**, assim como a espécie da qual esses alcaloides foram isolados.

**Tabela 8:** Lista dos 15 alcaloides com maiores potenciais de atividade previstos pelo consenso dos modelos LR-2D e SVM-3D.

| ID      | Alcaloides     | Estruturas em 2D                                                                                                     | pIC <sub>50</sub><br>previsto |
|---------|----------------|----------------------------------------------------------------------------------------------------------------------|-------------------------------|
| CAS1397 | Aspidolimidine |  <p><i>Apidosperma limae</i></p>   | 9,04                          |
| CAS1396 | Fendlerine     |  <p><i>Apidosperma limae</i></p>  | 8,84                          |
| CAS1394 | Aspidocarpine  |  <p><i>Apidosperma limae</i></p> | 8,74                          |

CAS1395

Aspidolimine

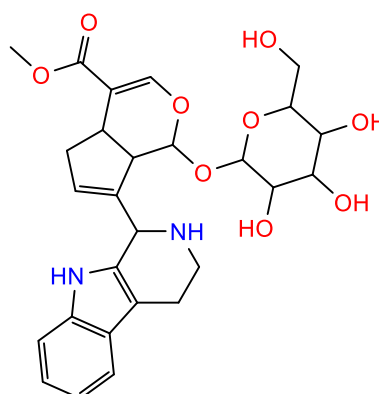


8,69

*Aspidosperma lima*

CAS1209

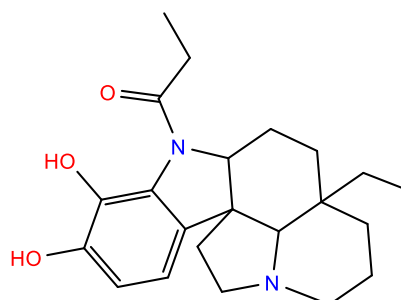
Psicolatina



8,68

*Psychotria (diferentes especies)*

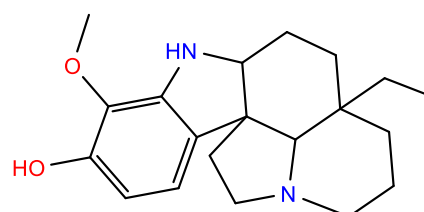
CAS911

N-propionil-  
16\_17-  
diidroxiaspidospe  
rmidina

8,45

*Aspidosperma melanocalyx*

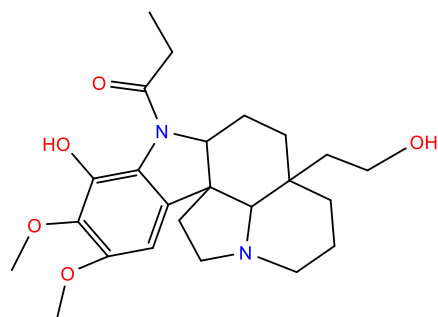
CAS918

16-  
desmetoxidesace  
tilpirifolidina

8,41

*Aspidosperma neblinae*

CAS920

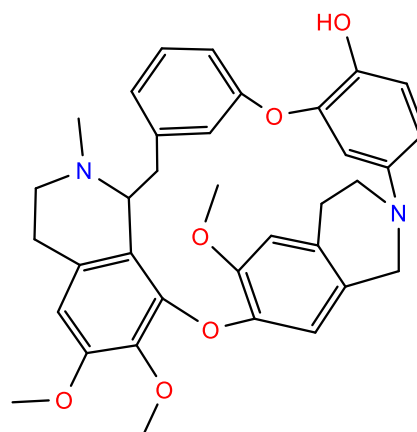
16-  
metoxilimapermin  
a

8,40

*Aspidosperma limae*

CAS1094

Sepeerine

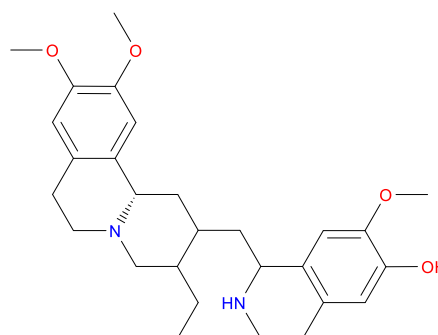


8,31

*Guatteria boliviana*

CAS1307

Cephaeline

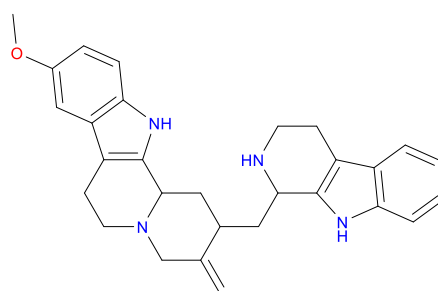


8,23

*Psychotria klugii*

CAS1295

Ramiflorines A

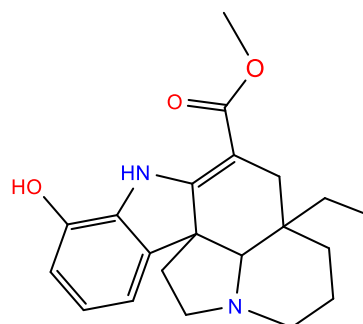


8,12

*Aspidosperma ramiflorum*

CAS909

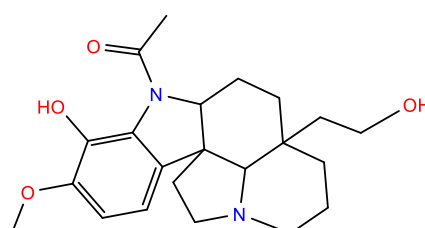
Vincadiformina



8,10

*Aspidosperma pyrifolium*

CAS924

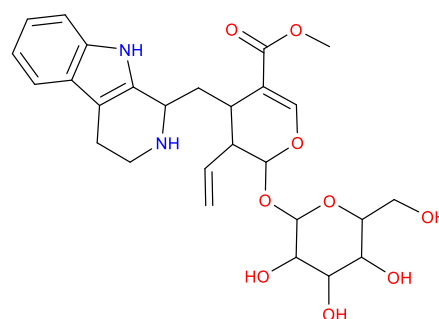
16-  
metoxilimapodina

8,10

*Aspidosperma limae*

CAS1416

Strictosidine

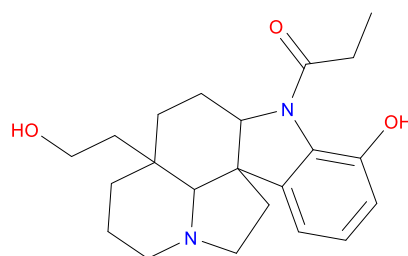


8,05

*Tabernaemontana littoralis*

CAS919

Limaspermina



8,02

*Aspidosperma limae*

Dos alcaloides selecionados como potencialmente ativos pelo consenso, o alcaloide ramiflorines A demonstrou conforme trabalho realizado por Tanaka et al., 2006, boa atividade sobre *S. aureus* [149]. Os demais alcaloides ainda não foram relatados na literatura como inibidores de *S. aureus* nem da enzima DHFR. De forma geral, acredita-se que os alcaloides selecionados das diferentes espécies de *Aspidosperma*, como aspidolimidine, fendlerine, aspidocarpine, aspidolimine, ainda que não avaliados isolados como inibidores de *S. aureus*, possam ser ativos, uma vez que estudos apontaram a atividade inibitória de extratos sobre *S. aureus* e bactérias gram-positivas [149–151].

Os alcaloides selecionados apresentam esqueletos de alcaloides isoquinolínicos, benzil e bisbenzilisquinolínicos, bem como alcaloides indolícos que representam classes de alcaloides que possuem atividade biológica definida para vários patógenos [7,14,77]. Assim, a triagem resultou em alcaloides com potencial atividade inibitória da SaDHFR e um estudo de *docking* molecular foi realizado na tentativa de elucidar o modo de ligação desses alcaloides na enzima SaDHFR.

### 5.3 *Docking* Molecular

O estudo do modo de ligação das moléculas de alcaloides selecionadas como ativas foi realizado com *docking* molecular dentro do sítio ativo da DHFR do *S. aureus* (código pdb 6DN2). O processo de seleção de parâmetros foi realizado através do *redocking* do ligante co-cristalizado que apresentou um RMSD de 0,162 usando a função de pontuação CHEMPLP em um raio de 12 Å. Todas as moléculas de água foram retiradas, pois em alguns casos essas moléculas podem interferir e blindar algumas interações importantes do ligante com a proteína [152].

Na **Tabela 9** encontram-se de forma resumida as interações intermoleculares dos alcaloides com os aminoácidos no sítio ativo SaDHFR obtidas pelo programa GOLD.

**Tabela 9:** Previsão do modo de ligação para os 15 alcaloides mais ativos obtidos pelo GOLD.

| <b>Alcaloides</b>                           | <b>Score</b> | <b>Interações com aminoácidos</b>                                                                                                            |
|---------------------------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Aspidolimidine                              | 53,70        | <b>Interações Hidrofóbicas:</b><br>Leu28, Val 31, Ile50, leu20, Xnp203<br><b>Interações Hidrogênio:</b> Ser49                                |
| Fendlerine                                  | 54,23        | <b>Interações Hidrofóbicas:</b><br>Phe92, Leu28, Val 31, Ile50, Leu20, Xnp203,<br><b>Interações Hidrogênio:</b> Ser49 2,37 Asp 27            |
| Aspidocarpine                               | 44,88        | <b>Interações Hidrofóbicas:</b><br>Phe92, Leu20, Leu28, His23, Val 31, Ile50, Leu20, Ser49                                                   |
| Aspidolimine                                | 47,22        | <b>Interações Hidrofóbicas:</b> Val 31, Lys52, Leu20, Xnp203, Asp 27                                                                         |
| Psicolatina                                 | 54,32        | <b>Interações Hidrofóbicas,</b> Val 31, Lys52, Leu20, Xnp203, Asp 27                                                                         |
| N-propionil-16_17-diidroxi-aspidospermidina | 52,43        | <b>Interações Hidrofóbicas:</b><br>Leu28, Ile50, Leu20,<br><b>Interações Hidrogênio:</b> His23                                               |
| 16-desmetoxidesacetilpirifolidina           | 45,30        | <b>Interações Hidrofóbicas:</b><br>Leu20, Ile50<br><b>Interações Hidrogênio:</b> Gln19                                                       |
| 16-metoxilimapermina                        | 44,52        | <b>Interações Hidrofóbicas:</b><br>Leu28,Leu28, Ile50, Phe92, Leu54,                                                                         |
| Sepeerine                                   | 46,10        | <b>Interações Hidrofóbicas:</b><br>Leu28, Ile50, Leu54,Lys32, Val31                                                                          |
| Cephaeline                                  | 55,88        | <b>Interações Hidrofóbicas:</b> His23, Leu28, Ile50, Lys29, Lys32, Val31, Leu54, Phe92, Xnp203<br><b>Interações Hidrogênio:</b> Asn18, Leu20 |
| Ramiflorines A                              | 58,46        | <b>Interações Hidrofóbicas:</b><br>Leu28, Ile50, Lys32, Leu54, Pro55, Gln19<br><b>Interações Hidrogênio:</b> Leu20                           |

|                     |       |                                                                                                                 |
|---------------------|-------|-----------------------------------------------------------------------------------------------------------------|
| Vincadiformina      | 49,29 | <b>Interações Hidrofóbicas:</b> Ala7, Leu28, Ile50, Leu54, Xnp203<br><b>Interações Hidrogênio:</b> Phe92, Gln19 |
| 16-metoxilimapodina | 49,45 | <b>Interações Hidrofóbicas:</b> Leu20, Leu28, Ile50, Leu54, Phe92, Val31                                        |
| Strictosidine       | 51,34 | <b>Interações Hidrofóbicas:</b> Leu20, Leu28, Leu54, His23, Pro55<br><b>Interações Hidrogênio:</b> Ser49, Asn18 |
| Limaspermina        | 52,14 | <b>Interações Hidrofóbicas:</b> Leu20, Leu28, Ile50, Xnp203, Val31, Phe92, His23                                |

A inspeção visual do *docking* (**Figura 30**) infere que as moléculas apresentam interações com aminoácidos importantes do sítio catalítico da DHFR. Ao analisar as melhores poses de cada molécula, foi possível observar as contribuições dos grupos hidroxila (OH), carbonila (C=O), grupo metoxi (OCH<sub>3</sub>) e do anel fenil para a formação de ligações de hidrogênio e hidrofóbicas.

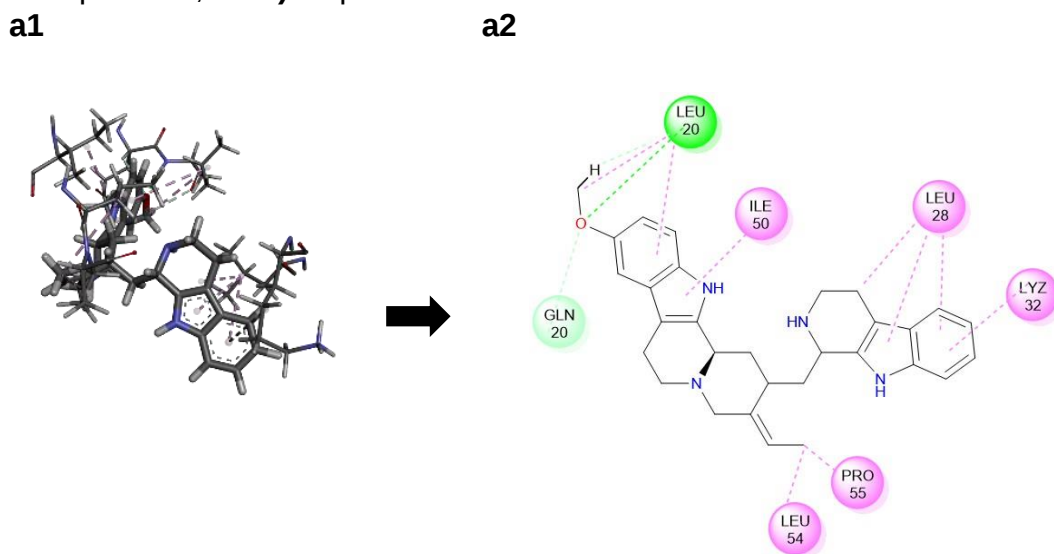
Entre os 15 alcaloides com maior potencial de inibição da DHFR, o alcaloide ramiflorines A, apresentou o maior valor de *Score*. Apesar da atividade do alcaloide ter sido relatada na literatura para *S. aureus*, o mecanismo de inibição não foi proposto. A ramiflorines A (**Figura 30, a1 e a2**), possui ligações de hidrogênio entre o resíduo Leu20 com o oxigênio do grupo metoxi. Além dessas interações, uma ligação de hidrogênio fraca (C-H) foi observada com o resíduo Gln19, esse tipo de ligação ocorre quando o átomo de C do ligante é o doador e o oxigênio da carbonila do aminoácido é o receptor [153]. Além disso, ligações hidrofóbicas importantes foram geradas com os aminoácidos Leu28, Ile50, Lys32, Leu54 e Pro55. Em particular, as interações entre DHFR e seus inibidores são predominantes hidrofóbicas [54].

Outros alcaloides também realizaram ligações hidrogênio com resíduos de aminoácidos. Os alcaloides aspidolimidine e fendlerine apresentaram ligações hidrogênio entre os resíduos Ser49 com oxigênios carbonílico (C=O) com comprimento de ligação de 2,30 e 2,37 Å, respectivamente. Estudo

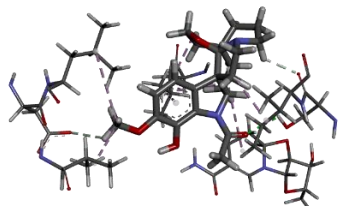
realizado por Ammar et al. (2021) identificou interações parecidas entre derivados de 7-etoxiquinolina e SaDHFR [48]. O alcaloide strictosidine gerou duas ligações de hidrogênio, uma de caráter doador e outra de caráter aceitador de hidrogênio entre o hidrogênio dos grupos hidroxila (O-H) com o aminoácido Asn18 a 1,97 Å e Ser49 a 2,13 Å, respectivamente. As interações polares entre a SaDHFR e os alcaloides com as interações entre oxigênio e o resíduo Ser49 são análogas ao encontrado pelo antifolato, TMP [54].

Os alcaloides também realizaram várias interações hidrofóbicas com diferentes aminoácidos do sítio ativo da SaDHFR. Dentre elas, interações alquila-arila (CH- $\pi$ ) e alquila são as mais frequentes. As interações CH- $\pi$  ocorrem quando um C-H sofre polarização das nuvens eletrônicas ocasionada por interações fracas do tipo dipolo-dipolo induzido na presença de um anel com sistema  $\pi$ . Ainda que fracas, essas interações são importantes para estabilidade de complexos proteína-ligante [154]. Foi verificado que todas as moléculas de alcaloides geraram esse tipo de interação com a enzima, devido aos anéis aromáticos presentes que induzem um momento dipolo nos grupos alquila dos aminoácidos. Na **Figura 30** pode ser visualizado as interações dos alcaloides melhores ranqueados.

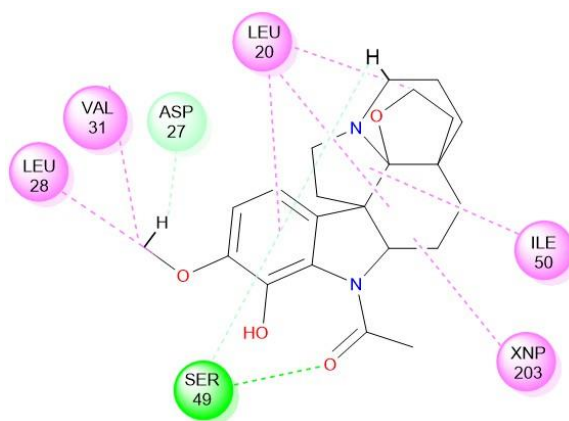
**Figura 30:** Estrutura 3D da melhor pose de cada alcaloide (à esquerda) e estrutura 2D com representação das interações (à direita). **a1/ a2)** Ramiflorines A, **b1/b2)** Aspidolimidine, **c1/c2)** Fendlerine, **d1/d2)** Strictosidine, **e1/e2)** Limaspermina, **f1/f2)** Cephaeline.



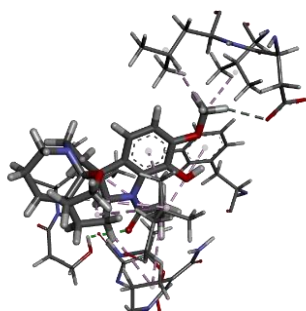
b1



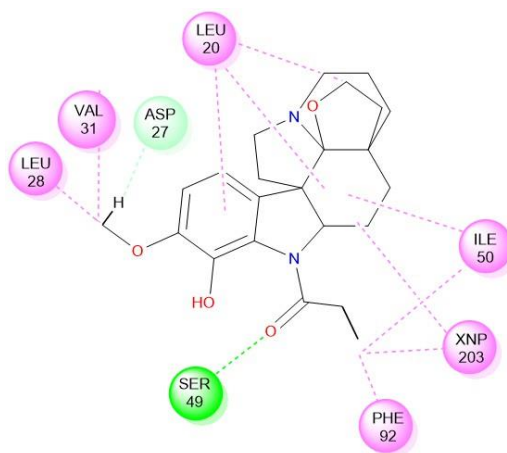
b2



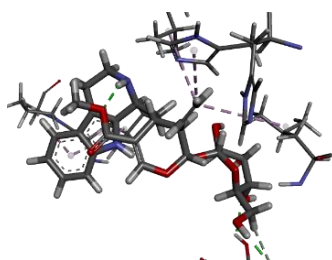
c1



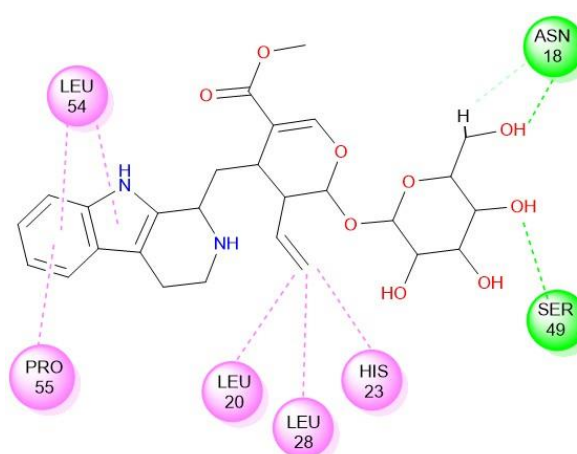
c2



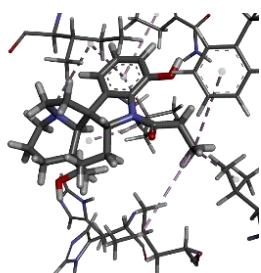
d1



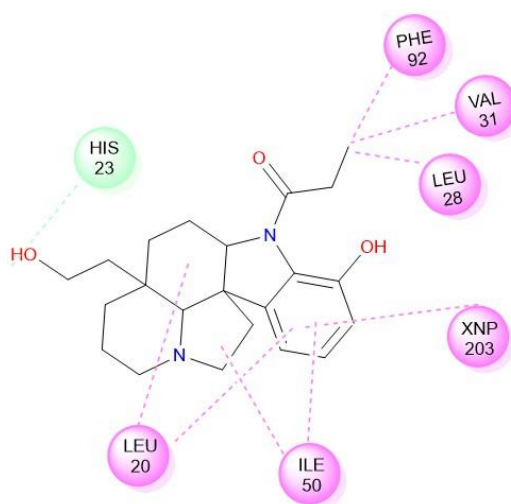
d2



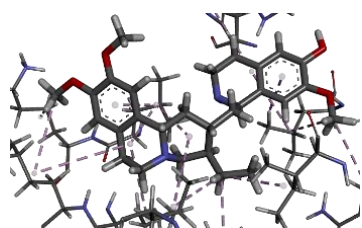
e1



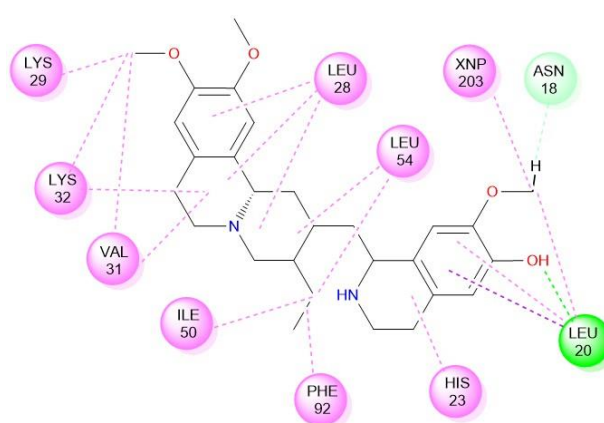
e2



f1



f2



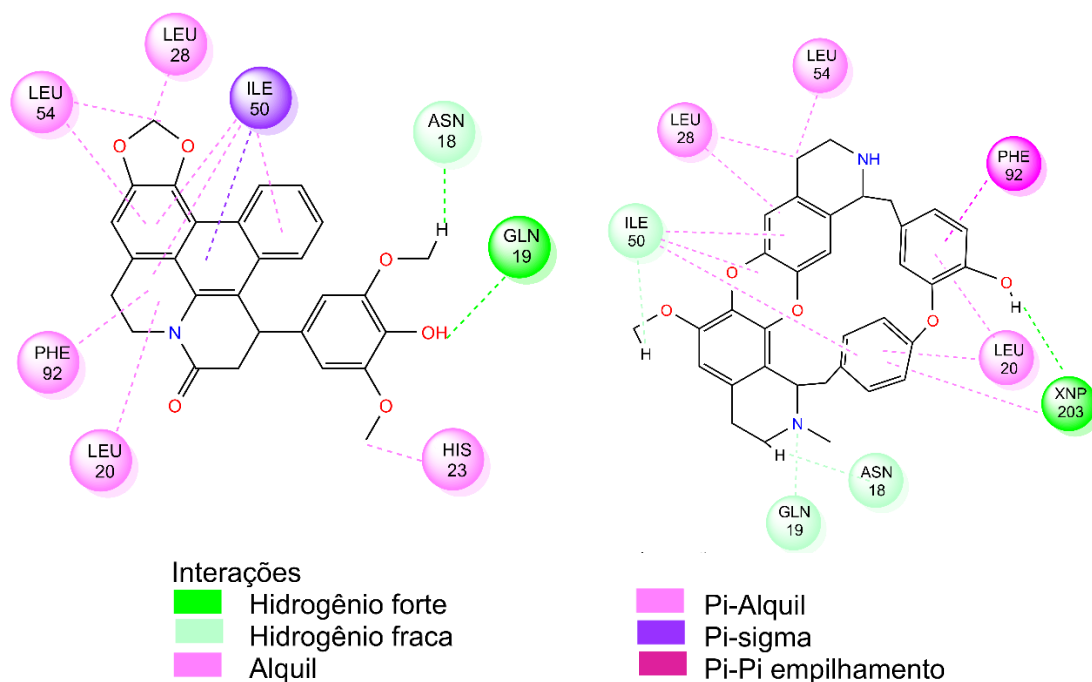
Fonte: próprio autor

Dos 30 alcaloides selecionados como ativos na triagem, a Aspateline e 7*H*-[1,3]Dioxolo[3,4]phenanthro[9,10,1-*ij*a]quinolizin-7-one, 5,6,9,10- tetrahydro-5- (4-hydroxy-3,5-dimethoxyphenyl) que chamaremos de QN7, mesmo não estando entre os 15 mais ativos, apresentaram os melhores scores, indicando que podem apresentar interações mais fortes com a SaDHFR. O primeiro alcaloide melhor ranqueado, QN7 (**Figura 31, (a)**) apresentou afinidade hidrofóbica do tipo alquila e alquila-arila com cinco resíduos de aminoácidos cujos os átomos estão dentro de 5 Å. Outra interação hidrofóbica foi visualizada para este alcaloide, a interação  $\pi$  - sigma que surgiu entre o anel aromático central do alcaloide com o resíduo Ile50. Essa interação não foi registrada para

nenhum dos 15 alcaloides mais ativos. Além das interações hidrofóbicas mencionadas, esse alcaloide apresenta duas interações de hidrogênio, H-O com o aminoácido Gln19 e C-H com a carbonila do resíduo Asn18.

O segundo alcaloide melhor ranqueado, Apateline (**Figura 31, (b)**), apresenta quatro interações de hidrogênio, sendo, uma interação forte entre O-H e o resíduo Xnp230 e três ligações fracas entre os resíduos Ile50, Gln19 e Asn18. Entre as ligações hidrofóbicas, alquila e alquila-arila com três resíduos de aminoácidos, o alcaloide apateline demonstra uma interação de empilhamento  $\pi$  entre o resíduo Phe92 e o anel fenil do fenol. Essa interação é um tipo de interação dipolo-dipolo entre anéis aromáticos, que ocorre devido à polarização das nuvens eletrônicas ou atração eletrostática entre os elétrons  $\pi$  do anel [155].

**Figura 31:** Estrutura 2D com representação das interações visualizadas para os dois alcaloides que apresentou os maiores valores scores. **a) QN7** e **b) Apateline**.



Fonte: próprio autor.

As interações de empilhamento  $\pi$  entre o anel aromático do alcaloide com o aminoácido Phe92, juntamente as interação  $\pi$ -sigma, são interações encontradas entre a SaDHFR e o antifolato iclaprim [156]. Tais interações

adicionais realizadas pelos alcaloides, QN7 e Apateline podem ser o motivo da maior afinidade observada entre esses alcaloides e o sítio ativo da enzima.

No geral, observa-se que todos os alcaloides interagem através de interações hidrofóbicas realizadas por elétrons  $\pi$ . As ligações  $\pi$  estão sujeitas a uma maior deformação da nuvem eletrônica, aumentando a intensidade do dipolo induzido, o que ocasionará um possível aumento na afinidade. Assim, foi possível inferir a partir da avaliação das interações do *docking* molecular, a importância do descritor SpMax\_Bh(p) que traz principalmente contribuições que podem explicar interações eletrostáticas das moléculas, CATS2D\_04\_DL e CATS3D\_03\_LL. Com isso, o presente estudo apresenta uma seleção eficiente para identificar propriedades físico-químicas importantes tanto para atividade biológica de alcaloides quanto para as interações no sítio catalítico SaDHFR.

## 6 CONCLUSÕES

Os métodos aplicados foram capazes de prever a atividade inibitória de estruturas químicas de alcaloides contra a enzima SaDHFR, pois os modelos QSAR construídos apresentaram resultados estatísticos condizentes com capacidade preditiva. O modelo MLP apresentou melhor desempenho na modelagem realizada para o banco de estruturas testadas contra o *S. aureus* e os modelos LR-2D e SVM-3D apresentaram melhores desempenhos na modelagem das estruturas testadas contra a SaDHFR. Deste modo, o modelo permitiu selecionar substâncias testadas puras e presentes em extratos de plantas com atividades contra *S. aureus*.

A partir de uma análise das contribuições dos descritores para atividade inibitória do *S. aureus*, foi inferida a importância das informações eletrônicas codificadas pelo descritor GATS5s, das ligações realizadas pelos pares [O-O] e [N-O], bem como contribuições dos fragmentos RNR2 e OH. Para a atividade de inibição de SaDHFR foi inferida a importância das características codificadas pela combinação dos descritores, SpMax3\_Bh(p), CATS2D\_04\_DL, CATS3D\_03\_LL e CATS3D\_05\_DL, o que sugere que a polarizabilidade das

moléculas e a presença de pares de átomos doadores de hidrogênio e lipofílicos devem favorecer o aumento na atividade inibitória.

Além disso, a partir do *docking* molecular, com a inspeção visual das interações intermoleculares entre os alcaloides e os aminoácidos de SaDHFR foram observadas interações de empilhamento  $\pi$ , interações hidrofóbicas e de hidrogênio. Esses resultados estão de acordo com as características fornecidas principalmente pelos descritores, GATS5s, SpMax3\_Bh(p), CATS2D\_04\_DL, CATS3D\_03\_LL e CATS3D\_05\_DL, que abrangem interações eletrônicas e estéricas das moléculas.

Assim, conclui-se que este estudo aplicou com sucesso diferentes abordagens *in silico*, resultando na identificação de 30 alcaloides potencialmente inibidores de SaDHFR, os quais apresentaram modo de ligação no sítio ativo da enzima que estão de acordo com resultados encontrados na literatura. Portanto, os alcaloides podem ser validados com estudos experimentais para direcionar o planejamento de novos candidatos a fármacos no tratamento de doenças infecciosas causadas por *S. aureus* em animais e humanos.

## **7 PERSPECTIVAS DO TRABALHO**

Os alcaloides previstos pelos modelos QSAR como inibidores de SaDHFR poderão ser comprovados e investigados experimentalmente com redução de custos financeiros e de animais usados em teste no desenvolvimento de novos candidatos a fármacos. Além disso, os resultados obtidos podem ser utilizados para prever outras substâncias que façam parte do mesmo espaço químico dos modelos.

## 8 REFERÊNCIAS

- [1] de Oliveira RP, Bezerra BA, de Melo RPB, da Silva DMS, de Carvalho RG, Juliano MA, et al. Bovine mastitis in northeastern Brazil: Occurrence of emergent bacteria and their phenotypic and genotypic profile of antimicrobial resistance. *Comp Immunol Microbiol Infect Dis* 2022;85:101802. <https://doi.org/10.1016/J.CIMID.2022.101802>.
- [2] Tomazi T, dos Santos MV. Antimicrobial use for treatment of clinical mastitis in dairy herds from Brazil and its association with herd-level descriptors. *Prev Vet Med* 2020;176:104937. <https://doi.org/10.1016/J.PREVETMED.2020.104937>.
- [3] Zhu C, Zhang M, Tang Q, Yang Q, Li J, He X, et al. Structure and Activity of the *Camellia oleifera* Sapogenin Derivatives on Growth and Biofilm Inhibition of *Staphylococcus aureus* and *Escherichia coli*. *J Agric Food Chem* 2019;67:14143–51. [https://doi.org/10.1021/ACS.JAFC.9B03577/ASSET/IMAGES/LARGE/JF9B03577\\_0008.JPEG](https://doi.org/10.1021/ACS.JAFC.9B03577/ASSET/IMAGES/LARGE/JF9B03577_0008.JPEG).
- [4] Cascioferro S, Raffa D, Maggio B, Raimondi MV, Schillaci D, Daidone G. Sortase A inhibitors: recent advances and future perspectives. *ACS Publications* 2015;58:9108–23. <https://doi.org/10.1021/acs.jmedchem.5b00779>.
- [5] Fergestad ME, Stamsås GA, Morales Angeles D, Salehian Z, Wasteson Y, Kjos M. Penicillin-binding protein PBP2a provides variable levels of protection toward different  $\beta$ -lactams in *Staphylococcus aureus* RN4220. *Microbiologyopen* 2020;9:e1057. <https://doi.org/10.1002/MBO3.1057>.
- [6] Liu C, Jeng W, Chang W, ... TK-J of B, 2012 undefined. Binding modes of zaragozic acid A to human squalene synthase and staphylococcal dehydrosqualene synthase. *ASBMB* 2012.
- [7] Singh G, Soni H, Tandon S, Kumar V, Babu G, Gupta V, et al. Identification of natural DHFR inhibitors in MRSA strains: Structure-based drug design study. *Results Chem* 2022;4:100292. <https://doi.org/10.1016/J.RECHEM.2022.100292>.

- [8] He J, Qiao W, An Q, Yang T, Luo Y. Dihydrofolate reductase inhibitors for use as antimicrobial agents. *Eur J Med Chem* 2020;195:112268. <https://doi.org/10.1016/J.EJMECH.2020.112268>.
- [9] Shamshad H, Bakri R, Mirza AZ. Dihydrofolate reductase, thymidylate synthase, and serine hydroxy methyltransferase: successful targets against some infectious diseases. *Mol Biol Rep* 2022;49:6659–91. <https://doi.org/10.1007/S11033-022-07266-8/FIGURES/4>.
- [10] Hoarau M, Vanichtanankul J, Srimongkolpithak N, Vitsupakorn D, Yuthavong Y, Kamchonwongpaisan S. Discovery of new non-pyrimidine scaffolds as *Plasmodium falciparum* DHFR inhibitors by fragment-based screening. *J Enzyme Inhib Med Chem* 2021;36:198–206. [https://doi.org/10.1080/14756366.2020.1854244/SUPPL\\_FILE/IENZ\\_A\\_1854244\\_SM1433.PDF](https://doi.org/10.1080/14756366.2020.1854244/SUPPL_FILE/IENZ_A_1854244_SM1433.PDF).
- [11] Heinberg A, Kirkman L. The molecular basis of antifolate resistance in *Plasmodium falciparum*: looking beyond point mutations. *Ann N Y Acad Sci* 2015;1342:10–8. <https://doi.org/10.1111/NYAS.12662>.
- [12] Cascioferro S, Carbone D, Parrino B, Pecoraro C, Giovannetti E, Cirrincione G, et al. Therapeutic strategies to counteract antibiotic resistance in MRSA biofilm-associated infections. *Wiley Online Library* 2021;16:65–80. <https://doi.org/10.1002/cmdc.202000677>.
- [13] Sharma M, Chauhan K, Shivahare R, Vishwakarma P, Suthar MK, Sharma A, et al. Discovery of a new class of natural product-inspired quinazolinone hybrid as potent antileishmanial agents. *J Med Chem* 2013;56:4374–92. [https://doi.org/10.1021/JM400053V/SUPPL\\_FILE/JM400053V\\_SI\\_001.PDF](https://doi.org/10.1021/JM400053V/SUPPL_FILE/JM400053V_SI_001.PDF).
- [14] Raimondi MV, Randazzo O, Franca M La, Barone G, Vignoni E, Rossi D, et al. DHFR inhibitors: Reading the past for discovering novel anticancer agents. *MdpiCom* 2019. <https://doi.org/10.3390/molecules24061140>.
- [15] Mesquita AA. Impacto da Mastite Bovina por *Staphylococcus aureus* e *Streptococcus agalactiae*. 2017.

- [16] IBGE. Produção Agropecuária no Brasil | IBGE 2022. <https://www.ibge.gov.br/explica/producao-agropecuaria/> (accessed November 26, 2022).
- [17] Chen X, Chen Y, Zhang W, Chen S, Wen X, Ran X, et al. Prevalence of subclinical mastitis among dairy cattle and associated risks factors in China during 2012–2021: A systematic review and meta-analysis. *Res Vet Sci* 2022;148:65–73. <https://doi.org/10.1016/J.RVSC.2022.04.007>.
- [18] He W, Ma S, Lei L, He J, Li X, Tao J, et al. Prevalence, etiology, and economic impact of clinical mastitis on large dairy farms in China. *Vet Microbiol* 2020;242:108570. <https://doi.org/10.1016/J.VETMIC.2019.108570>.
- [19] Pegolo S, Tessari R, Bisutti V, Vanzin A, Giannuzzi D, Giancesella M, et al. Quarter-level analyses of the associations among subclinical intramammary infection and milk quality, udder health, and cheesemaking traits in Holstein cows. *J Dairy Sci* 2022;105:3490–507. <https://doi.org/10.3168/JDS.2021-21267>.
- [20] Coatrini-Soares A, Coatrini-Soares J, Popolin Neto M, Mello SS de, Pinto DDSC, Carvalho WA, et al. Microfluidic E-tongue to diagnose bovine mastitis with milk samples using Machine learning with Decision Tree models. *Chemical Engineering Journal* 2022;451:138523. <https://doi.org/10.1016/J.CEJ.2022.138523>.
- [21] Mesquita AA, Rocha CMBM, Bruhn FRP, Custódio DAC, Braz MS, Pinto SM, et al. *Staphylococcus aureus* and *Streptococcus agalactiae*: prevalence, resistance to antimicrobials, and their relationship with the milk quality of dairy cattle herds in Minas Gerais state, Brazil. *Pesquisa Veterinária Brasileira* 2019;39:308–16. <https://doi.org/10.1590/1678-5150-PVB-5821>.
- [22] Bradley AJ. Bovine Mastitis: An Evolving Disease. *The Veterinary Journal* 2002;164:116–28. <https://doi.org/10.1053/TVJL.2002.0724>.
- [23] Demontier E, Dubé-Duquette A, Brouillette E, Larose A, Ster C, Lucier JF, et al. Relative virulence of *Staphylococcus aureus* bovine mastitis strains representing the main Canadian spa types and clonal complexes as

- determined using in vitro and in vivo mastitis models. *J Dairy Sci* 2021;104:11904–21. <https://doi.org/10.3168/JDS.2020-19904>.
- [24] Alves VM, Braga RC, Muratov EN, Andrade CH. Cheminformatics: An introduction. *Quim Nova* 2018;41:202–12. <https://doi.org/10.21577/0100-4042.20170145>.
- [25] Roshan M, Parmanand, Arora D, Behera M, Vats A, Gautam D, et al. Virulence and enterotoxin gene profile of methicillin-resistant *Staphylococcus aureus* isolates from bovine mastitis. *Comp Immunol Microbiol Infect Dis* 2022;80:101724. <https://doi.org/10.1016/J.CIMID.2021.101724>.
- [26] Chang BS, Moon JS, Kang HM, Kim YI, Lee HK, Kim JD, et al. Protective effects of recombinant staphylococcal enterotoxin type C mutant vaccine against experimental bovine infection by a strain of *Staphylococcus aureus* isolated from subclinical mastitis in dairy cattle. *Vaccine* 2008;26:2081–91. <https://doi.org/10.1016/j.vaccine.2008.02.043>.
- [27] Dittmann KK, Chaul LT, Lee SHI, Corassin CH, de Oliveira CAF, De Martinis ECP, et al. *Staphylococcus aureus* in some Brazilian dairy industries: Changes of contamination and diversity. *Front Microbiol* 2017;8:2049. <https://doi.org/10.3389/FMICB.2017.02049/BIBTEX>.
- [28] Dos Santos AL, Santos DO, De Freitas CC, Ferreira BLA, Afonso IF, Rodrigues CR, et al. *Staphylococcus aureus*: visitando uma cepa de importância hospitalar. *J Bras Patol Med Lab* 2007;43:413–23. <https://doi.org/10.1590/S1676-24442007000600005>.
- [29] Tuchscher LPN, Gomez MI, Buzzola FR, Calvinho LF, Lee JC, Sordelli DO. Characterization of a new variant of IS257 that has displaced the capsule genes within bovine isolates of *Staphylococcus aureus*. *Infect Immun* 2007;75:5483–8. <https://doi.org/10.1128/IAI.00747-07/ASSET/7B759948-FD65-4FC6-A422-CBBDE521A5FD/ASSETS/GRAPHIC/ZII0110770050003.JPEG>.
- [30] Benito D, Gómez P, Aspiroz C, Zarazaga M, Lozano C, Torres C. Molecular characterization of *Staphylococcus aureus* isolated from humans related to a livestock farm in Spain, with detection of MRSA-CC130 carrying *mecC* gene:

- A zoonotic case? *Enferm Infecc Microbiol Clin* 2016;34:280–5.  
<https://doi.org/10.1016/J.EIMC.2015.03.008>.
- [31] Jaiswal R, Garg A, Tripathi P, Venkatesh V. Epidemiology of Pantone Valentine Leukocidin in clinical *Staphylococcus aureus* isolates - A prospective study at a tertiary care centre in North India. *Clin Epidemiol Glob Health* 2022;15:101006. <https://doi.org/10.1016/J.CEGH.2022.101006>.
- [32] Iwatsuki K, Yamasaki O, Morizane S, Oono T. Staphylococcal cutaneous infections: Invasion, evasion and aggression. *J Dermatol Sci* 2006;42:203–14.  
<https://doi.org/10.1016/J.JDERMSCI.2006.03.011>.
- [33] Rossi BF, Bonsaglia ECR, Pantoja JCF, Santos M V., Gonçalves JL, Fernandes Júnior A, et al. Short communication: Association between the accessory gene regulator (agr) group and the severity of bovine mastitis caused by *Staphylococcus aureus*. *J Dairy Sci* 2021;104:3564–8.  
<https://doi.org/10.3168/jds.2020-19275>.
- [34] Sacco SC, Velázquez NS, Renna MS, Beccaria C, Baravalle C, Pereyra EAL, et al. Capacity of two *Staphylococcus aureus* strains with different adaptation genotypes to persist and induce damage in bovine mammary epithelial cells and to activate macrophages. *Microb Pathog* 2020;142:104017.  
<https://doi.org/10.1016/J.MICPATH.2020.104017>.
- [35] Molineri AI, Camussone C, Zbrun MV, Suárez Archilla G, Cristiani M, Neder V, et al. Antimicrobial resistance of *Staphylococcus aureus* isolated from bovine mastitis: Systematic review and meta-analysis. *Prev Vet Med* 2021;188:105261. <https://doi.org/10.1016/J.PREVETMED.2021.105261>.
- [36] Moreira LH, de Souza JCP, de Lima CJ, Salgado MAC, Fernandes AB, Andreani DIK, et al. Use of photodynamic therapy in the treatment of bovine subclinical mastitis. *Photodiagnosis Photodyn Ther* 2018;21:246–51.  
<https://doi.org/10.1016/J.PDPDT.2017.12.009>.
- [37] Freu G, Tomazi T, Filho AFS, Heinemann MB, Dos Santos M V. Antimicrobial Resistance and Molecular Characterization of *Staphylococcus aureus* Recovered from Cows with Clinical Mastitis in Dairy Herds from Southeastern

- Brazil. *Antibiotics* 2022, Vol 11, Page 424 2022;11:424.  
<https://doi.org/10.3390/ANTIBIOTICS11040424>.
- [38] Rao S, Linke L, Magnuson R, Jaunch L, Hyatt DR. Antimicrobial resistance and genetic diversity of *Staphylococcus aureus* collected from livestock, poultry and humans. *One Health* 2022;15:100407.  
<https://doi.org/10.1016/J.ONEHLT.2022.100407>.
- [39] Foster TJ. Antibiotic resistance in *Staphylococcus aureus*. Current status and future prospects. *FEMS Microbiol Rev* 2017;41:430–49.  
<https://doi.org/10.1093/FEMSRE/FUX007>.
- [40] Naranjo-Lucena A, Slowey R. Invited review: Antimicrobial resistance in bovine mastitis pathogens: A review of genetic determinants and prevalence of resistance in European countries. *J Dairy Sci* 2022.  
<https://doi.org/10.3168/jds.2022-22267>.
- [41] Abdeen EE, Mousa WS, Abdel Salam SY, Al-Maary KS, Mubarak AS, Moussa IM, et al. Antibigram and phylogenetic diversity of enterotoxigenic *Staphylococcus aureus* strains from milk products and public health implications. *Saudi J Biol Sci* 2020;27:1968–74.  
<https://doi.org/10.1016/J.SJBS.2020.06.028>.
- [42] Holmes MA, Zadoks RN. Methicillin resistant *S. aureus* in human and bovine mastitis. *J Mammary Gland Biol Neoplasia* 2011;16:373–82.  
<https://doi.org/10.1007/S10911-011-9237-X/FIGURES/1>.
- [43] Sanad SMH, Mekky AEM, El-Idreesy TT. Potential bacterial biofilm, MRSA, and DHFR inhibitors based on new morpholine-linked chromene-thiazole hybrids: One-pot synthesis and in silico study. *J Mol Struct* 2022;1248:131476.  
<https://doi.org/10.1016/J.MOLSTRUC.2021.131476>.
- [44] Gonen N, Assaraf YG. Antifolates in cancer therapy: Structure, activity and mechanisms of drug resistance. *Drug Resistance Updates* 2012;15:183–210.  
<https://doi.org/10.1016/J.DRUP.2012.07.002>.
- [45] Jansen G, Weijers K, Blits M, Van Der Heijden J, Gent Y, Al M, et al. Folates and antifolates in rheumatoid arthritis. *Pteridines* 2013;24:21–6.

<https://doi.org/10.1515/PTERID-2013-0012/MACHINEREADABLECITATION/RIS>.

- [46] Azzam RA, Elsayed RE, Elgemeie GH. Design, Synthesis, and Antimicrobial Evaluation of a New Series of N-Sulfonamide 2-Pyridones as Dual Inhibitors of DHPS and DHFR Enzymes. *ACS Omega* 2020;5:10401–14. [https://doi.org/10.1021/ACSOMEGA.0C00280/SUPPL\\_FILE/AO0C00280\\_SI\\_001.PDF](https://doi.org/10.1021/ACSOMEGA.0C00280/SUPPL_FILE/AO0C00280_SI_001.PDF).
- [47] Lombardo MN, Dayanandan NG, Keshipeddy S, Zhou W, Si D, Reeve SM, et al. Structure-guided in vitro to in vivo pharmacokinetic optimization of propargyl-linked antifolates. *Drug Metabolism and Disposition* 2019;47:995–1003. <https://doi.org/10.1124/dmd.119.086504>.
- [48] Ammar YA, El-Hafez SMAA, Hessein SA, Ali AM, Askar AA, Ragab A. One-pot strategy for thiazole tethered 7-ethoxy quinoline hybrids: Synthesis and potential antimicrobial agents as dihydrofolate reductase (DHFR) inhibitors with molecular docking study. *J Mol Struct* 2021;1242:130748. <https://doi.org/10.1016/J.MOLSTRUC.2021.130748>.
- [49] Reeve SM, Si D, Krucinska J, Yan Y, Viswanathan K, Wang S, et al. Toward Broad Spectrum Dihydrofolate Reductase Inhibitors Targeting Trimethoprim Resistant Enzymes Identified in Clinical Isolates of Methicillin Resistant *Staphylococcus aureus*. *ACS Infect Dis* 2019;5:1896–906. [https://doi.org/10.1021/ACSINFECDIS.9B00222/SUPPL\\_FILE/ID9B00222\\_SI\\_001.PDB](https://doi.org/10.1021/ACSINFECDIS.9B00222/SUPPL_FILE/ID9B00222_SI_001.PDB).
- [50] Wróbel A, Arciszewska K, Maliszewski D, Drozdowska D. Trimethoprim and other nonclassical antifolates an excellent template for searching modifications of dihydrofolate reductase enzyme inhibitors. *J Antibiot (Tokyo)* 2019;73:5–27. <https://doi.org/10.1038/S41429-019-0240-6>.
- [51] Li X, Hilgers M, Cunningham M, Chen Z, Trzoss M, Zhang J, et al. Structure-based design of new DHFR-based antibacterial agents: 7-aryl-2,4-diaminoquinazolines. *Bioorg Med Chem Lett* 2011;21:5171–6. <https://doi.org/10.1016/J.BMCL.2011.07.059>.

- [52] Li X, Hilgers M, Cunningham M, Chen Z, Trzoss M, Zhang J, et al. Structure-based design of new DHFR-based antibacterial agents: 7-aryl-2,4-diaminoquinazolines. *Bioorg Med Chem Lett* 2011;21:5171–6. <https://doi.org/10.1016/J.BMCL.2011.07.059>.
- [53] Dale GE, Broger C, D'Arcy A, Hartman PG, DeHoogt R, Jolidon S, et al. A single amino acid substitution in *Staphylococcus aureus* dihydrofolate reductase determines trimethoprim resistance. *J Mol Biol* 1997;266:23–30. <https://doi.org/10.1006/JMBI.1996.0770>.
- [54] Muddala NP, White JC, Nammalwar B, Pratt I, Thomas LM, Bunce RA, et al. Inhibitor design to target a unique feature in the folate pocket of *Staphylococcus aureus* dihydrofolate reductase. *Eur J Med Chem* 2020;200:112412. <https://doi.org/10.1016/J.EJMECH.2020.112412>.
- [55] Peppard WJ, Schuenke CD. Iclaprim, a diaminopyrimidine dihydrofolate reductase inhibitor for the potential treatment of antibiotic-resistant staphylococcal infections. *Curr Opin Investig Drugs* 2008;9:210–25.
- [56] Rodrigues RP, Mantoani SP, De Almeida JR, Pinsetta FR, Semighini EP, Da Silva VB, et al. Virtual screening strategies in drug design. *Revista Virtual de Quimica* 2012;4:739–76. <https://doi.org/10.5935/1984-6835.20120055>.
- [57] Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-based virtual screening: Advances and applications in drug discovery. *Front Pharmacol* 2018;9:1275. <https://doi.org/10.3389/FPHAR.2018.01275/BIBTEX>.
- [58] Willett P. Chemoinformatics: a history. *Wiley Interdiscip Rev Comput Mol Sci* 2011;1:46–56. <https://doi.org/10.1002/WCMS.1>.
- [59] Gillet VJ, Leach AR. Chemoinformatics. *Comprehensive Medicinal Chemistry II* 2020;3:635–76. <https://doi.org/10.1016/B978-0-12-409547-2.14327-6>.
- [60] Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* 1962 194:4824 1962;194:178–80. <https://doi.org/10.1038/194178b0>.

- [61] Muratov EN, Bajorath J, Sheridan RP, Tetko I V., Filimonov D, Poroikov V, et al. QSAR without borders. *Chem Soc Rev* 2020;49:3525–64. <https://doi.org/10.1039/D0CS00098A>.
- [62] Gramatica P. Principles of QSAR Modeling. *International Journal of Quantitative Structure-Property Relationships* 2020;5:61–97. <https://doi.org/10.4018/IJQSPR.20200701.OA1>.
- [63] Portugal I, Alencar P, Cowan D. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Syst Appl* 2018;97:205–27. <https://doi.org/10.1016/J.ESWA.2017.12.020>.
- [64] Chen B, Zhang T, Bond T, Gan Y. Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: A review of methods and resources. *J Hazard Mater* 2015;299:260–79. <https://doi.org/10.1016/j.jhazmat.2015.06.054>.
- [65] Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Ravichandran Veerasamy, et al : Validation of QSAR Models-Strategies and Importance Validation of QSAR Models-Strategies and Importance. n.d.
- [66] Wu Z, Zhu M, Kang Y, Leung ELH, Lei T, Shen C, et al. Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets. *Brief Bioinform* 2021;22. <https://doi.org/10.1093/bib/bbaa321>.
- [67] Muhammad U, Uzairu A, Arthur DE. REVIEW ON: QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR) MODELING. vol. 4. 2018.
- [68] Király P, Kiss R, Kovács D, Ballaj A, Tóth G. The Relevance of Goodness-of-fit, Robustness and Prediction Validation Categories of OECD-QSAR Principles with Respect to Sample Size and Model Type. *Mol Inform* 2022;41:2200072. <https://doi.org/10.1002/MINF.202200072>.
- [69] Mukai M, Nagao K, Yamaguchi S, Ohmiya H. Molecular Field Analysis Using Computational-Screening Data in Asymmetric N-Heterocyclic Carbene-Copper Catalysis toward Data-Driven In Silico Catalyst Optimization.

- <https://doi.org/10.1246/Bcsj20210349> 2022; 95:271–7.  
<https://doi.org/10.1246/BCSJ.20210349>.
- [70] Damale MG, Harke SN, Kalam Khan FA, Shinde DB, Sangshetti JN. Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. vol. 14. 2014.
- [71] Sousa SF, Ribeiro AJM, Coimbra JTS, Neves RPP, Martins SA, Moorthy NSHN, et al. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Curr Med Chem* 2013; 20:2296–314. <https://doi.org/10.2174/0929867311320180002>.
- [72] Jain AN, Nicholls A. Recommendations for evaluation of computational methods. *J Comput Aided Mol Des* 2008; 22:133–9. <https://doi.org/10.1007/S10822-008-9196-5/METRICS>.
- [73] Asiamah I, Obiri SA, Tamekloe W, Armah FA, Borquaye LS. Applications of molecular docking in natural products-based drug discovery. *Sci Afr* 2023;20. <https://doi.org/10.1016/j.sciaf.2023.e01593>.
- [74] El-Gazzar YI, Georgey HH, El-Messery SM, Ewida HA, Hassan GS, Raafat MM, et al. Synthesis, biological evaluation and molecular modeling study of new (1,2,4-triazole or 1,3,4-thiadiazole)-methylthio-derivatives of quinazolin-4(3H)-one as DHFR inhibitors. *Bioorg Chem* 2017; 72:282–92. <https://doi.org/10.1016/J.BIOORG.2017.04.019>.
- [75] Kobayashi M, Kinjo T, Koseki Y, Bourne CR, Barrow WW, Aoki S. Identification of novel potential antibiotics against *Staphylococcus* using structure-based drug screening targeting dihydrofolate reductase. *J Chem Inf Model* 2014;54:1242–53. <https://doi.org/10.1021/ci400686d>.
- [76] Matos I de A, Goes Pinto AC, Ferraz MVF, Adan WCS, Rodrigues RP, dos Santos JX, et al. Identification of potential *Staphylococcus aureus* dihydrofolate reductase inhibitors using QSAR, molecular docking, dynamics simulations and free energy calculation. *J Biomol Struct Dyn* 2022. [https://doi.org/10.1080/07391102.2022.2057361/SUPPL\\_FILE/TBSD\\_A\\_2057361\\_SM4638.DOCX](https://doi.org/10.1080/07391102.2022.2057361/SUPPL_FILE/TBSD_A_2057361_SM4638.DOCX).

- [77] Rahman S, Das AK. Integrated Multi-omics, Virtual Screening and Molecular Docking Analysis of Methicillin-Resistant *Staphylococcus aureus* USA300 for the Identification of Potential Therapeutic Targets: An In-Silico Approach. *International Journal of Peptide Research and Therapeutics* 2021 27:4 2021;27:2735–55. <https://doi.org/10.1007/S10989-021-10287-9>.
- [78] ChEMBL Database n.d. <https://www.ebi.ac.uk/chembl/> (accessed September 28, 2023).
- [79] Fourches D, Muratov E, Tropsha A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* 2010; 50:1189–204. [https://doi.org/10.1021/CI100176X/ASSET/IMAGES/LARGE/CI-2010-00176X\\_0009.JPEG](https://doi.org/10.1021/CI100176X/ASSET/IMAGES/LARGE/CI-2010-00176X_0009.JPEG).
- [80] Fourches D, Muratov E, Tropsha A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J Chem Inf Model* 2016;56:1243–52. [https://doi.org/10.1021/ACS.JCIM.6B00129/ASSET/IMAGES/LARGE/CI-2016-00129X\\_0005.JPEG](https://doi.org/10.1021/ACS.JCIM.6B00129/ASSET/IMAGES/LARGE/CI-2016-00129X_0005.JPEG).
- [81] Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010; 29:476–88. <https://doi.org/10.1002/minf.201000061>.
- [82] The RDKit Documentation — The RDKit 2023.03.1 documentation n.d. <https://rdkit.org/docs/index.html> (accessed September 10, 2023).
- [83] Stewart Computational Chemistry - MOPAC Home Page n.d. <http://openmopac.net/> (accessed September 10, 2023).
- [84] Sun Q, Gieseeking RLM. Parametrization of the PM7 Semiempirical Quantum Mechanical Method for Silver Nanoclusters. *Journal of Physical Chemistry A* 2022. [https://doi.org/10.1021/ACS.JPCA.2C05782/SUPPL\\_FILE/JP2C05782\\_SI\\_002.ZIP](https://doi.org/10.1021/ACS.JPCA.2C05782/SUPPL_FILE/JP2C05782_SI_002.ZIP).
- [85] alvaDesc - Alvascience. Alvascience n.d. <https://www.alvascience.com/alvadesec/> (accessed September 10, 2023).

- [86] Hewitt M, Cronin MTD, Madden JC, Rowe PH, Johnson C, Obi A, et al. Consensus QSAR models: Do the benefits outweigh the complexity? *J Chem Inf Model* 2007; 47:1460–8. [https://doi.org/10.1021/CI700016D/SUPPL\\_FILE/CI700016D-FILE002.PDF](https://doi.org/10.1021/CI700016D/SUPPL_FILE/CI700016D-FILE002.PDF).
- [87] Liu H, Papa E, Gramatica P. QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chem Res Toxicol* 2006; 19:1540–8. [https://doi.org/10.1021/TX0601509/SUPPL\\_FILE/TX0601509SI20060703\\_120023.PDF](https://doi.org/10.1021/TX0601509/SUPPL_FILE/TX0601509SI20060703_120023.PDF).
- [88] Rácz A, Bajusz D, Héberger K. Intercorrelation Limits in Molecular Descriptor Preselection for QSAR/QSPR. *Mol Inform* 2019; 38:1800154. <https://doi.org/10.1002/MINF.201800154>.
- [89] Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. *J Big Data* 2020; 7. <https://doi.org/10.1186/s40537-020-00327-4>.
- [90] Rasyid H, Hariani Soekamto N, Seniwati, Firdausiah S, Firdaus. Modelling the Anticancer Activity of 4-Alkoxy Cinnamic Analogues using 3D-Descriptors and Genetic Algorithm-Multiple Linear Regression (GA-MLR) Method. *J King Saud Univ Sci* 2023;35:102514. <https://doi.org/10.1016/J.JKSUS.2022.102514>.
- [91] Kurniawan I, Tarwidi D, Jondri. QSAR modeling of PTP1B inhibitor by using Genetic algorithm-Neural network methods. *J Phys Conf Ser*, vol. 1192, Institute of Physics Publishing; 2019. <https://doi.org/10.1088/1742-6596/1192/1/012059>.
- [92] Roy K, Kar S, Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems* 2015;145:22–9. <https://doi.org/10.1016/j.chemolab.2015.04.013>.
- [93] Alipour MM, Razavi SN, Feizi Derakhshi MR, Balafar MA. A hybrid algorithm using a genetic algorithm and multiagent reinforcement learning heuristic to solve the traveling salesman problem. *Neural Comput Appl* 2018;30:2935–51. <https://doi.org/10.1007/s00521-017-2880-4>.

- [94] Gramatica P. Principles of QSAR models validation: Internal and external. *QSAR Comb Sci* 2007;26:694–701. <https://doi.org/10.1002/qsar.200610151>.
- [95] Driouche Y, Messadi D. Quantitative structure-retention relationship model for predicting retention indices of constituents of essential oils of *Thymus vulgaris* (Lamiaceae). *Journal of the Serbian Chemical Society* 2019; 84:405–16. <https://doi.org/10.2298/JSC180817010D>.
- [96] Weka 3 - Data Mining with Open Source Machine Learning Software in Java n.d. <https://www.cs.waikato.ac.nz/ml/weka/> (accessed September 28, 2023).
- [97] Baumann K. Cross-validation as the objective function for variable-selection techniques. *TrAC Trends in Analytical Chemistry* 2003; 22:395–406. [https://doi.org/10.1016/S0165-9936\(03\)00607-1](https://doi.org/10.1016/S0165-9936(03)00607-1).
- [98] Golbraikh A, Tropsha A. QSAR modeling using chirality descriptors derived from molecular topology. *J Chem Inf Comput Sci* 2003; 43:144–54. <https://doi.org/10.1021/CI025516B/ASSET/IMAGES/LARGE/CI025516BF00004.JPEG>.
- [99] Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Igglessi-Markopoulou O, Kollias G. A combined LS-SVM & MLR QSAR workflow for predicting the inhibition of CXCR3 receptor by quinazolinone analogs. *Mol Divers* 2010;14:225–35. <https://doi.org/10.1007/s11030-009-9163-7>.
- [100] Melagraki G, Afantitis A, Sarimveis H, Koutentis PA, Kollias G, Igglessi-Markopoulou O. Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. *Mol Divers* 2009; 13:301–11. <https://doi.org/10.1007/s11030-009-9115-2>.
- [101] Rücker C, Rücker G, Meringer M. Y-randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 2007; 47:2345–57. <https://doi.org/10.1021/CI700157B/ASSET/IMAGES/LARGE/CI700157BF00002.JPEG>.
- [102] Khan MF, Verma G, Akhtar W, Shaquiquzzaman M, Akhter M, Rizvi MA, et al. Pharmacophore modeling, 3D-QSAR, docking study and ADME prediction of acyl 1,3,4-thiadiazole amides and sulfonamides as antitubulin agents. *Arabian*

- Journal of Chemistry 2019; 12:5000–18.  
<https://doi.org/10.1016/J.ARABJC.2016.11.004>.
- [103] Zhang S, Golbraikh A, Oloff S, Kohn H, Tropsha A. A novel Automated Lazy Learning QSAR (ALL-QSAR) approach: Method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. J Chem Inf Model 2006; 46:1984–95.  
<https://doi.org/10.1021/ci060132x>.
- [104] Paixão V V M. Estudo *In Silico* para a Seleção de Alcaloides com Potencial Atividade de Inibição aa Tio redoxina Glutathione Redutase oo *Schistosoma mansoni*. 2022.
- [105] CAS SciFinder. CAS SciFinder<sup>n</sup> | CAS. <https://www-cas.ez20.periodicos.capes.gov.br/solutions/cas-scifinder-discovery-platform/cas-scifinder> (accessed October 26, 2022).
- [106] RCSB PDB: Homepage n.d. <https://www.rcsb.org/> (accessed September 28, 2023).
- [107] Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today 2016; 21:1291–302.  
<https://doi.org/10.1016/J.DRUDIS.2016.06.013>.
- [108] Du Z, Wang D, Li Y. Comprehensive Evaluation and Comparison of Machine Learning Methods in QSAR Modeling of Antioxidant Tripeptides. ACS Omega 2022;7:25760–71.  
[https://doi.org/10.1021/ACSOMEGA.2C03062/SUPPL\\_FILE/AO2C03062\\_SI\\_002.ZIP](https://doi.org/10.1021/ACSOMEGA.2C03062/SUPPL_FILE/AO2C03062_SI_002.ZIP).
- [109] Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today 2016; 21:1291–302.  
<https://doi.org/10.1016/J.DRUDIS.2016.06.013>.
- [110] Ibrahim MT, Uzairu A, Uba S, Shallangwa GA. Computational modeling of novel quinazoline derivatives as potent epidermal growth factor receptor inhibitors. Heliyon 2020; 6:e03289.  
<https://doi.org/10.1016/J.HELİYON.2020.E03289>.

- [111] Beheshti A, Pourbasheer E, Nekoei M, Vahdani S. QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions. *Journal of Saudi Chemical Society* 2016; 20:282–90. <https://doi.org/10.1016/j.jscs.2012.07.019>.
- [112] Golbraikh A, Tropsha A. Beware of q<sup>2</sup>! *J Mol Graph Model* 2002; 20:269–76. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1).
- [113] Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 2003;111:1361–75. <https://doi.org/10.1289/EHP.5758>.
- [114] Sepehri B, Kohnepoushi M, Ghavami R. High predictive QSAR models for predicting the SARS coronavirus main protease inhibition activity of ketone-based covalent inhibitors. *Journal of the Iranian Chemical Society* 2022;19:1865–76. <https://doi.org/10.1007/s13738-021-02426-2>.
- [115] Golbraikh A, Bonchev D, Tropsha A. Novel Chirality Descriptors Derived from Molecular Topology. *J Chem Inf Comput Sci* 2001; 41:147–58. <https://doi.org/10.1021/CI000082A/ASSET/IMAGES/MEDIUM/CI000082AU00002A.GIF>.
- [116] Ricci-Lopez J, Aguila SA, Gilson MK, Brizuela CA. Improving Structure-Based Virtual Screening with Ensemble Docking and Machine Learning. *J Chem Inf Model* 2021; 61:5362–76. [https://doi.org/10.1021/ACS.JCIM.1C00511/ASSET/IMAGES/LARGE/CI1C00511\\_0006.JPEG](https://doi.org/10.1021/ACS.JCIM.1C00511/ASSET/IMAGES/LARGE/CI1C00511_0006.JPEG).
- [117] Martins JPA, Ferreira MMC. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. *Quim Nova* 2013; 36:554–60. <https://doi.org/10.1590/S0100-40422013000400013>.
- [118] Shen M, Béguin C, Golbraikh A, Stables JP, Kohn H, Tropsha A. Application of Predictive QSAR Models to Database Mining: Identification and Experimental Validation of Novel Anticonvulsant Compounds. *J Med Chem* 2004;47:2356–64. [https://doi.org/10.1021/JM030584Q/SUPPL\\_FILE/JM030584Q\\_S.PDF](https://doi.org/10.1021/JM030584Q/SUPPL_FILE/JM030584Q_S.PDF).

- [119] Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des* 2020; 34:1013–26. <https://doi.org/10.1007/s10822-020-00314-0>.
- [120] Zhong S, Guan X. Developing Quantitative Structure–Activity Relationship (QSAR) Models for Water Contaminants’ Activities/Properties by Fine-Tuning GPT-3 Models. *Environ Sci Technol Lett* 2023. <https://doi.org/10.1021/acs.estlett.3c00599>.
- [121] Kier LB, Hall LH, Frazer JW. AN INDEX OF ELECTROTOPOLOGICAL STATE FOR ATOMS IN MOLECULES. vol. 7. 1991.
- [122] Pogliani L. Modeling with molecular pseudoconnectivity descriptors. A useful extension of the intrinsic I-state concent. *Journal of Physical Chemistry A* 2000;104:9029–45. <https://doi.org/10.1021/JP001191V/ASSET/IMAGES/MEDIUM/JP001191VU00015A.GIF>.
- [123] Velázquez-Libera JL, Caballero J, Toropova AP, Toropov AA. Estimation of 2D autocorrelation descriptors and 2D Monte Carlo descriptors as a tool to build up predictive models for acetylcholinesterase (AChE) inhibitory activity. *Chemometrics and Intelligent Laboratory Systems* 2019; 184:14–21. <https://doi.org/10.1016/J.CHEMOLAB.2018.11.008>.
- [124] Geary RC. The Contiguity Ratio and Statistical Mapping. vol. 5. 1954.
- [125] Roy J, Ghosh S, Ojha PK, Roy K. Predictive quantitative structure-property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs). *Environ Sci Nano* 2019; 6:224–47. <https://doi.org/10.1039/c8en01059e>.
- [126] Hall LH, Kier LB. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. vol. 35. 1995.
- [127] Contrera JF, Matthews EJ, Benz RD. Predicting the carcinogenic potential of pharmaceuticals in rodents using molecular structural similarity and E-state

- indices. *Regulatory Toxicology and Pharmacology* 2003; 38:243–59. [https://doi.org/10.1016/S0273-2300\(03\)00071-0](https://doi.org/10.1016/S0273-2300(03)00071-0).
- [128] Pan Y, Jiang J, Wang R, Cao H, Zhao J. Prediction of auto-ignition temperatures of hydrocarbons by neural network based on atom-type electrotopological-state indices. *J Hazard Mater* 2008; 157:510–7. <https://doi.org/10.1016/J.JHAZMAT.2008.01.016>.
- [129] Todeschini R and C V. *Molecular Descriptors Chemoinformatics*. vol. 1, 2009, p. 107–11. <https://doi.org/10.1002/9783527628766.fmatter1>.
- [130] De Julián-Ortiz J V., De Gregorio Alapont C, Ríos-Santamarina I, García-Doménech R, Gálvez J. Prediction of properties of chiral compounds by molecular topology. *J Mol Graph Model* 1998; 16:14–8. [https://doi.org/10.1016/S1093-3263\(98\)00013-8](https://doi.org/10.1016/S1093-3263(98)00013-8).
- [131] Schneider N, Lewis RA, Fechner N, Ertl P. Chiral Cliffs: Investigating the Influence of Chirality on Binding Affinity. *ChemMedChem* 2018; 13:1315–24. <https://doi.org/10.1002/cmdc.201700798>.
- [132] Dragos H, Gilles M, Alexandre V. Predicting the predictability: A unified approach to the applicability domain problem of qsar models. *J Chem Inf Model* 2009; 49:1762–76. [https://doi.org/10.1021/CI9000579/SUPPL\\_FILE/CI9000579\\_SI\\_001.ZIP](https://doi.org/10.1021/CI9000579/SUPPL_FILE/CI9000579_SI_001.ZIP).
- [133] Yu X. Prediction of Depuration Rate Constants for Polychlorinated Biphenyl Congeners. *ACS Omega* 2019; 4:15615–20. [https://doi.org/10.1021/ACSOMEGA.9B02072/ASSET/IMAGES/MEDIUM/AO9B02072\\_M010.GIF](https://doi.org/10.1021/ACSOMEGA.9B02072/ASSET/IMAGES/MEDIUM/AO9B02072_M010.GIF).
- [134] Seribeli FL, Maximiano FA. Interações Intermoleculares: conteúdos conceituais sobre o tema na graduação em química nas opiniões de pesquisadores. *Revista Da Sociedade Brasileira de Ensino de Química* 2021;2:e022108. <https://doi.org/10.56117/resbenq.2021.v2.e022108>.
- [135] Procópio MB, Cesar A. Polarizabilidades e primeira hiperpolarizabilidades elétricas dipolares de ésteres E-4-amino-trans-1,3-butadienil-(1',2'-di-

- hidroxibenzeno) de boro, alumínio e gálio. *Quim Nova* 2009; 32:1157–63. <https://doi.org/10.1590/S0100-40422009000500014>.
- [136] Arroio A, Honório KM, Da Silva ABF. Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade. *Quim Nova* 2010;33:694–9. <https://doi.org/10.1590/S0100-40422010000300037>.
- [137] Ciura K, Ulenberg S, Kapica H, Kawczak P, Belka M, Bączek T. Drug affinity to human serum albumin prediction by retention of cetyltrimethylammonium bromide pseudostationary phase in micellar electrokinetic chromatography and chemically advanced template search descriptors. *J Pharm Biomed Anal* 2020;188:113423. <https://doi.org/10.1016/J.JPBA.2020.113423>.
- [138] Ghosh A, Panda P, Halder AK, Cordeiro MNDS. In silico characterization of aryl benzoyl hydrazide derivatives as potential inhibitors of RdRp enzyme of H5N1 influenza virus. *Front Pharmacol* 2022; 13:1004255. <https://doi.org/10.3389/FPHAR.2022.1004255/BIBTEX>.
- [139] Fechner U, Franke L, Renner S, Schneider P, Schneider G. Comparison of correlation vector methods for ligand-based similarity searching. vol. 17. 2003.
- [140] Neugebauer A, Hartmann RW, Klein CD. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *J Med Chem* 2007;50:4665–8. <https://doi.org/10.1021/jm070533j>.
- [141] Randic M. Molecular Shape Profiles1. vol. 35. 1995.
- [142] Asadollahi-Baboli M. Quantitative structure-activity relationship analysis of human neutrophil elastase inhibitors using shuffling classification and regression trees and adaptive neuro-fuzzy inference systems. *SAR QSAR Environ Res* 2012; 23:505–20. <https://doi.org/10.1080/1062936X.2012.665811>.
- [143] Grisoni F, Merk D, Consonni V, Hiss JA, Tagliabue SG, Todeschini R, et al. Scaffold hopping from natural products to synthetic mimetics by holistic molecular similarity. *Communications Chemistry* 2018 1:1 2018; 1:1–9. <https://doi.org/10.1038/s42004-018-0043-x>.

- [144] Grisoni F, Merk D, Byrne R, Schneider G. Scaffold-Hopping from Synthetic Drugs by Holistic Molecular Representation. *Sci Rep* 2018; 8. <https://doi.org/10.1038/s41598-018-34677-0>.
- [145] Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V, Todeschini R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* 2012; 17:4791–810. <https://doi.org/10.3390/molecules17054791>.
- [146] Marzo M, Kulkarni S, Manganaro A, Roncaglioni A, Wu S, Barton-Maclaren TS, et al. Integrating in silico models to enhance predictivity for developmental toxicity. *Toxicology* 2016; 370:127–37. <https://doi.org/10.1016/J.TOX.2016.09.015>.
- [147] Valsecchi C, Grisoni F, Consonni V, Ballabio D. Consensus versus Individual QSARs in Classification: Comparison on a Large-Scale Case Study. *J Chem Inf Model* 2020; 60:1215–23. [https://doi.org/10.1021/ACS.JCIM.9B01057/ASSET/IMAGES/LARGE/CI9B01057\\_0002.JPEG](https://doi.org/10.1021/ACS.JCIM.9B01057/ASSET/IMAGES/LARGE/CI9B01057_0002.JPEG).
- [148] Abdelaziz A, Spahn-Langguth H, Schramm KW, Tetko I V. Consensus modeling for HTS assays using in silico descriptors calculates the best balanced accuracy in Tox21 challenge. *Front Environ Sci* 2016; 4. <https://doi.org/10.3389/fenvs.2016.00002>.
- [149] Tanaka JCA, Da Silva CC, De Oliveira AJB, Nakamura C V, Dias Filho BP. Antibacterial activity of indole alkaloids from *Aspidosperma ramiflorum*. vol. 39. 2006.
- [150] De Oliveira AJB, Koike L, Reis FDAMH, Eguchi SY, Endo EH, Nakamura CV, et al. Preliminary studies on the antibacterial activity of crude extracts and alkaloids from species of *Aspidosperma*. *Pharm Biol* 2009; 47:1085–9. <https://doi.org/10.3109/13880200903008633>.
- [151] Pessini G.L. APGV, BVB, CMA, NCV, REAN, SAEG, A-JJX 3,. Evaluation of antimicrobial activity of three *Aspidosperma* species. *Special ISSUE* 2012;1:112–9.

- [152] Roberts BC, Mancera RL. Ligand - protein docking with water molecules. *J Chem Inf Model* 2008;48:397–408. <https://doi.org/10.1021/ci700285e>.
- [153] Kumar M, Balaji P V. C-H...pi interactions in proteins: Prevalence, pattern of occurrence, residue propensities, location, and contribution to protein stability. *J Mol Model* 2014;20:1–14. <https://doi.org/10.1007/S00894-014-2136-5/METRICS>.
- [154] Fokoue HH, Pinheiro PSM, Fraga CAM, Sant'Anna CMR. Is there anything new about the molecular recognition applied to medicinal chemistry? *Quim Nova* 2020; 43:78–89. <https://doi.org/10.21577/0100-4042.20170474>.
- [155] Salonen LM, Ellermann M, Diederich F. Aromatic rings in chemical and biological recognition: energetics and structures. *Angew Chem Int Ed Engl* 2011;50:4808–42. <https://doi.org/10.1002/ANIE.201007560>.
- [156] Oefner C, Bandera M, Haldimann A, Laue H, Schulz H, Mukhija S, et al. Increased hydrophobic interactions of iclaprim with *Staphylococcus aureus* dihydrofolate reductase are responsible for the increase in affinity and antibacterial activity. *Journal of Antimicrobial Chemotherapy* 2009; 63:687–98. <https://doi.org/10.1093/JAC/DKP024>.

## 9 APÊNDICES

### 9.1 Apêndice 1- Descritores selecionados para construção dos modelos

**Tabela 10:** Lista dos descritores selecionados pelo método AG para desenvolver os modelos (conjunto SaDB).

| Classe                        | Descritores 2D | Definição Específica                                     | Definição Geral                                                                                                                                                                                                          | Ref. |
|-------------------------------|----------------|----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| Autocorrelações 2D            | GATS5s         | Autocorrelação de Geary de lag 5 ponderada pelo estado I | Os descritores de autocorrelação codificam a posição relativa dos átomos considerando uma propriedade distribuída ao longo de uma distância topológica, em termos de números de ligações (2D) ou distâncias euclidianas. | [1]  |
| Contagem de grupos funcionais | nRNR2          | Número de aminas terciárias (alifáticas)                 | Os descritores de contagem de grupos funcionais codificam informações químicas importantes de grupos bem conhecidos.                                                                                                     | [2]  |
|                               | nN(CO)2        | Número de imidas                                         |                                                                                                                                                                                                                          |      |
|                               | nArCOOH        | Número de ácidos carboxílicos                            |                                                                                                                                                                                                                          |      |

|                                   |          |                                                           |                                                                                                                                                                                                                                               |       |
|-----------------------------------|----------|-----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                                   |          | aromáticos                                                |                                                                                                                                                                                                                                               |       |
|                                   | nOHp     | Número de álcool secundário                               |                                                                                                                                                                                                                                               |       |
|                                   | nOHt     | Número de álcool terciário                                |                                                                                                                                                                                                                                               |       |
| Descritores de Anel               | nR07     | Número de anéis de 7 membros                              | São descritores que quantificam a presença de anéis em uma molécula.                                                                                                                                                                          | [3]   |
| Pares de átomos 2D                | F04[N-O] | 1. Frequência / ausência de O-O na distância topológica 6 | Os descritores de pares de átomos, calcula a frequência, ausência ou presença de pares de átomos separados por um caminho mais curto. O grafo de distância é definido como o menor número de átomos ao longo de um caminho entre dois átomos. | [4]   |
|                                   | B06[O-O] | Presença / ausência de O-O na distância topológica 6      |                                                                                                                                                                                                                                               |       |
|                                   | B09[O-O] | Presença / ausência de O-O na distância topológica 9      |                                                                                                                                                                                                                                               |       |
| Índices de estado E do tipo átomo | minsCH3  | Minimum CH3                                               | Esse descritor combina informações eletrônicas e topológicas do átomo analisado, estado de valência, características atômicas,                                                                                                                | [5,6] |

|                                   |              |                                                                                       |                                                                                                                                                                                     |        |
|-----------------------------------|--------------|---------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
|                                   |              |                                                                                       | acessibilidade de elétrons. Descrevi moléculas no nível atômico.                                                                                                                    |        |
| Descritores de carga              | qpmax        | Carga positiva máxima dos átomos em uma molécula                                      | É usado para descrever características eletrônicas de toda a molécula e de regiões particulares, como de átomos, ligações e fragmentos.                                             | [7]    |
| Desscritores de quiralidade       | s3_phSize    | Número de átomos de N/O do substituinte 4 normalizado pelos átomos N/O compartilhados | E desenvolvidos com base em informações topológicas de grafos moleculares que incluem índices de conectividade molecular, índices de grupo de Zagreb e índices de carga topológica. | [8]    |
| Descritores baseados em matriz 2D | SpMax_X      | Autovalor principal da matriz chi                                                     | São descritores de autovalor relacionados a matriz de distâncias topológicas que capturam informações tanto da conectividade quanto informações atômicas de uma molécula.           | [9,10] |
| Descritores de farmacóforo        | CATS2D_03_DP | CATS2D Doador-Positivo no lag 03                                                      | Os descritores <i>CATS</i> , pertencem ao grupo de descritores de vetores de                                                                                                        | [11]   |

|                               |        |                                                                                   |                                                                                                                                                                |      |
|-------------------------------|--------|-----------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
|                               |        |                                                                                   | correlação que codificam informações sobre frequências de pares de átomos, levando em consideração distâncias topológicas (2D) ou distâncias euclidianas (3D). |      |
| Fragmentos centrados no átomo | C-18   | =CHX                                                                              | Os descritores de fragmentos centrados no átomo trazem informações de átomos de hidrogênio, átomos de carbono, e heteroátomos.                                 | [7]  |
| Propriedades moleculares      | ALOGP2 | Coeficiente de partição octanol-água de Ghose-Crippen ao quadrado. ( $\log P^2$ ) | São descritores que codificam informações farmacocinéticas da molécula.                                                                                        | [12] |

**Tabela 11:** Lista dos descritores selecionados pelo método AG para desenvolver os modelos (conjunto SadhfrDB).

| Classe | Descritores 2D | Definição Específica | Definição Geral | Ref. |
|--------|----------------|----------------------|-----------------|------|
|--------|----------------|----------------------|-----------------|------|

|                                   |              |                                                                                                          |                                                                                                                                                                                                                                   |        |
|-----------------------------------|--------------|----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Descritores baseados em matriz 2D | SpMax3_Bh(p) | Máximo autovalor absoluto da matriz modificada de Burden – n3 ponderada pela polarizabilidades relativas | São descritores de autovalor relacionados a matriz de distâncias topológicas que capturam informações tanto da conectividade quanto i nformações atômicas de uma molécula.                                                        | [9,10] |
|                                   | SpMax_X      | Autovalor principal da matriz chi                                                                        |                                                                                                                                                                                                                                   |        |
|                                   | SpMin1_Bh(m) | Mínimo autovalor absoluto da matriz modificada de Burdem – n1 ponderada pela massa                       |                                                                                                                                                                                                                                   |        |
| Descritores CATS3D                | CATS2D_04_DL | CATS3D Doador-Lipofílico BIN 04 (4.000 – 5.000 Å)                                                        | Os descritores CATS, pertencem ao grupo de descritores de vetores de correlação que codificam informações sobre frequências de pares de átomos, levando em consideração distâncias tologicas (2D) ou diatâncias euclidianas (3D). | [11]   |

| Classe             | Descritores 3D | Definição Expecífica                       | Definição Geral                                                                   | Ref. |
|--------------------|----------------|--------------------------------------------|-----------------------------------------------------------------------------------|------|
| Perfil moleculares | SHP2           | índice médio de perfil de forma de ordem 2 | Leva em consideração a média de perfil de vários descritores inferiores derivadas | [13] |

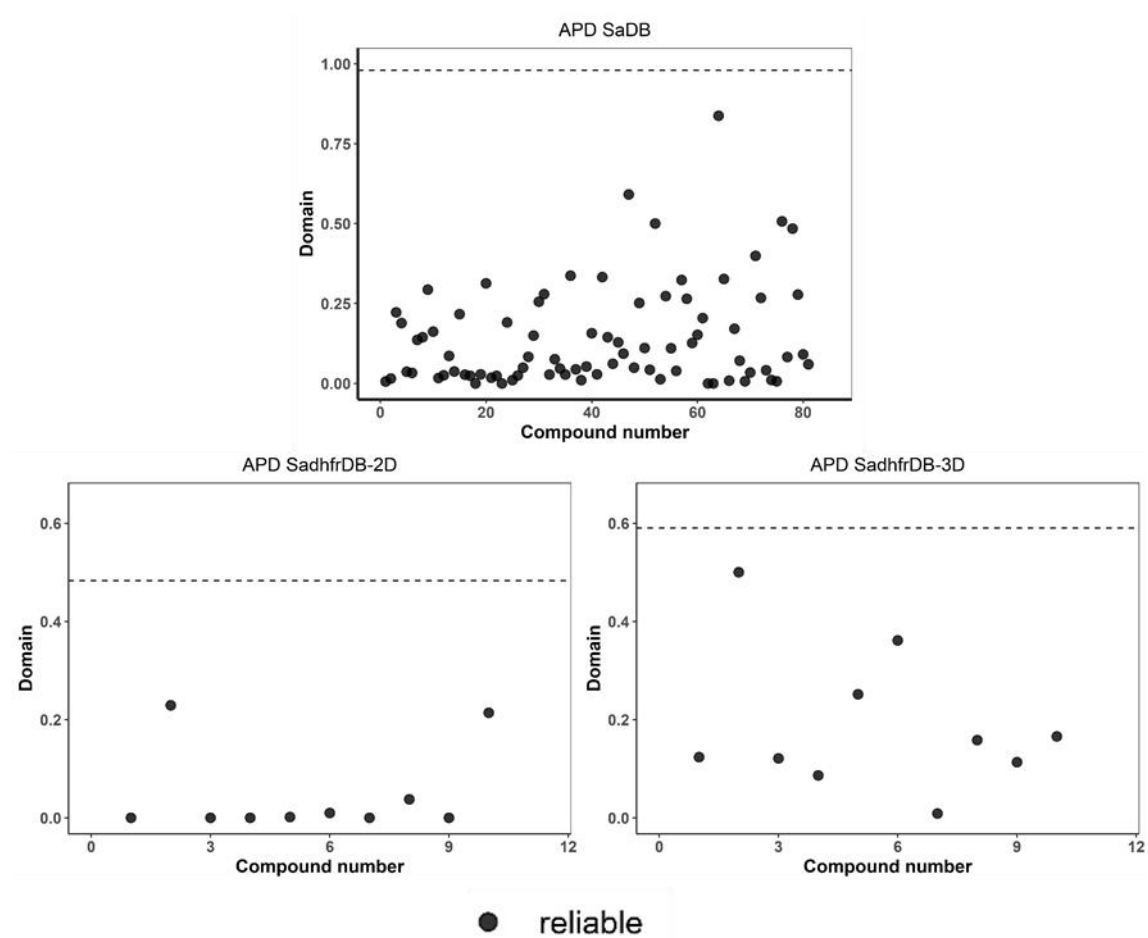
|                        |               |                                                         |                                                                                                                                                                                                                                    |         |
|------------------------|---------------|---------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| de Randic              |               |                                                         | da distância interatômicas dos átomos na periferia de uma molécula. São baseados na teoria dos grafos e envolve no cálculo várias características das moleculares como forma molecular, tamanho e extensão.                        |         |
| CATS3D                 | CATS3D_05_DL  | CATS3D Doador-Lipofílico BIN 05 (5.000 – 6.000 Å)       | Os descritores CATS, pertencem ao grupo de descritores de vetores de correlação que codificam informações sobre frequências de pares de átomos, levando em consideração distâncias topológicas (2D) ou distâncias euclidianas (3D) | [11]    |
|                        | CATS3D_03_LL  | CATS3D Lipofílico - Lipofílico BIN 03 (3.000 – 4.000 Å) |                                                                                                                                                                                                                                    |         |
|                        | CATS3D_08_LL  | CATS3D Lipofílico - Lipofílico BIN 08 (8.000 – 9.000 Å) |                                                                                                                                                                                                                                    |         |
| Descritores de WHALES* | WHALES100_Rem | WHALES Remoteness (Rem) (percentile 100)                | Os descritores WHALES , capturam informações de distâncias geométricas de interação entre os átomos, formato molecular e distribuição de carga parcial de forma holística.                                                         | [14,15] |

\*WHALES- *Weighttted Holistic Atom Localization and Entity Shape*

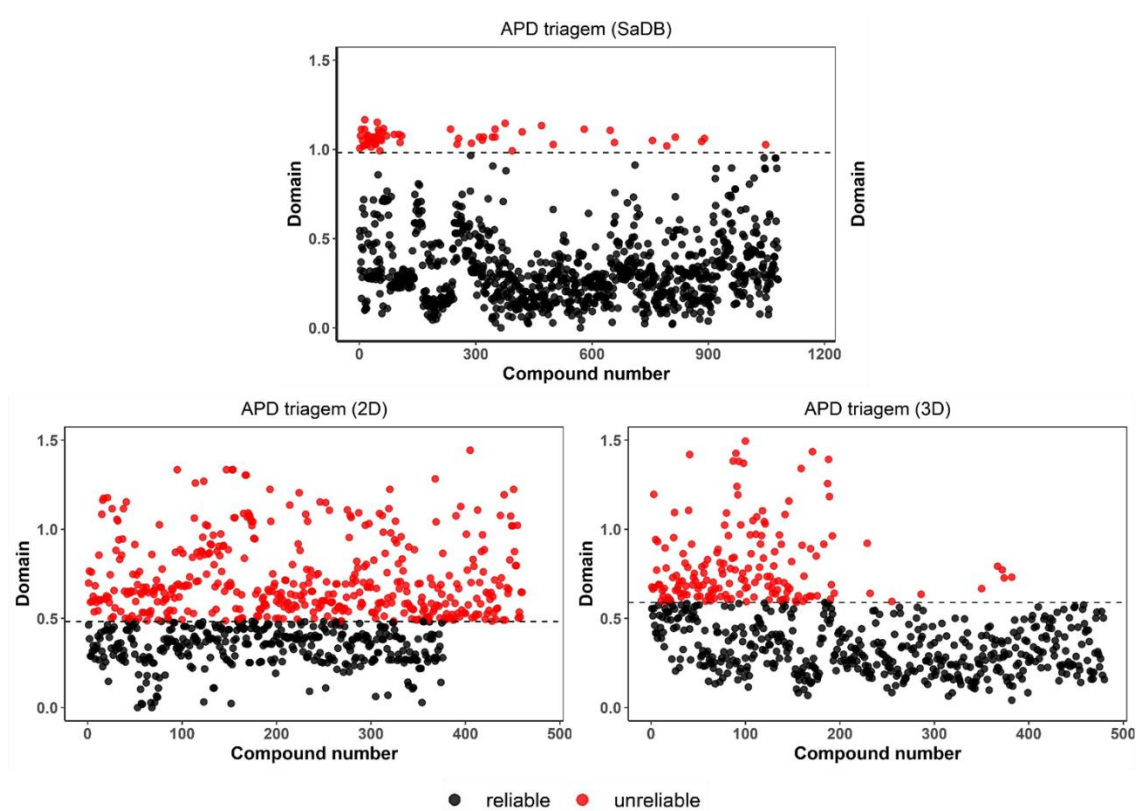
Fonte: próprio autor

## 9.2 Apêndice 2- Domínio de Aplicabilidade

**Figura 32:** Domínio de aplicabilidade dos compostos do conjunto teste baseados nas distâncias euclidianas.



**Figura 33:** Dominio de aplicabilidade dos compostos do conjunto de triagem baseados nas distâncias euclidianas.



#### 9.4 Apêndice 3- Seleção dos alcaloides potencialmente ativos pelo consenso da triagem virtual nos modelos LR-2D e SVM-3D.

**Tabela 12:** Previsão da atividade dos alcaloides na SaDHFR.

| <b>CAS number</b> | <b>Alcaloides</b>                          | <b>Classificação</b>         | <b>pIC<sub>50</sub></b> |
|-------------------|--------------------------------------------|------------------------------|-------------------------|
| CAS1397           | Aspidolimidine                             | Indólico                     | 9,04                    |
| CAS1396           | Fendlerine                                 | Indólico                     | 8,84                    |
| CAS1394           | Aspidocarpine                              | Indólico                     | 8,74                    |
| CAS1395           | Aspidolimine                               | Indólico                     | 8,69                    |
| CAS1209           | Psicolatina                                | Indólico                     | 8,68                    |
| CAS914            | Aspidolimina                               | Indólico                     | 8,58                    |
| CAS911            | N-propionil-16_17-diidroxiaspidospermidina | Indólico                     | 8,45                    |
| CAS918            | 16-Desmetoxidesacetilpirifolidina          | Indólico                     | 8,41                    |
| CAS920            | 16-Metoxilimapermina                       | Indólico                     | 8,40                    |
| CAS1094           | Sepeerine                                  | Bisbenzylisoquinoline        | 8,31                    |
| CAS1307           | Cephaeline                                 | Benzoquinolizidine           | 8,23                    |
| CAS1295           | Ramiflorines A                             | Indólico monoterpênico       | 8,12                    |
| CAS909            | Vincadiformina                             | Indólico                     | 8,10                    |
| CAS924            | 16-Metoxilimapodina                        | Indólico                     | 8,10                    |
| CAS1416           | strictosidine                              | Indólico monoterpênico       | 8,05                    |
| CAS919            | Limaspermina                               | Indólico                     | 8,02                    |
| CAS1137           | Guattaminone                               | Não Informado                | 8,02                    |
| CAS925            | N-formicilindrocarpinol                    | Indólico                     | 7,94                    |
| CAS663            | (S) N-methylcoclaurine                     | Benzyltetrahydroisoquinoline | 7,93                    |
| CAS1419           | Tabersonine                                | Indólicos monoterpênicos     | 7,90                    |
| CAS1092           | Tiliageine                                 | Bisbenzylisoquinoline        | 7,90                    |
| CAS1361           | Lagerine                                   | Bifenilquinolizidina         | 7,85                    |

|         |                                                                                                                              |                       |      |
|---------|------------------------------------------------------------------------------------------------------------------------------|-----------------------|------|
| CAS992  | 15_16_17-Trimetoxi-21-oxoaspidoalbina                                                                                        | Indólico              | 7,80 |
| CAS1214 | Epi-des-7-O-methylroraimine                                                                                                  | Bisbenzylisoquinoline | 7,80 |
| CAS1359 | 5-epi-dihydrolyfoline                                                                                                        | Bifenilquinolizidina  | 7,79 |
| CAS1178 | 7H-[1,3]Dioxolo[3,4]phenanthro[9,10,1- <i>ija</i> ]quinolizin-7-one, 5,6,9,10- tetrahydro-5- (4-hydroxy-3,5-dimethoxyphenyl) | Nonisoquinoline       | 7,72 |
| CAS1306 | Kluginine                                                                                                                    | Benzoquinolizidine    | 7,71 |
| CAS1101 | Apateline                                                                                                                    | Arpofinico            | 7,64 |

## 9.5 Apêndice 4- Referências das tabelas 11 e 12

- [1] Sliwoski G, Mendenhall J, Meiler J. Autocorrelation descriptor improvements for QSAR: 2DA-Sign and 3DA-Sign. *J Comput Aided Mol Des* 2016;30:209–17. <https://doi.org/10.1007/s10822-015-9893-9>.
- [2] Pirovano A, Brandmaier S, Huijbregts MAJ, Ragas AMJ, Veltman K, Hendriks AJ. The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals. *Environ Toxicol Pharmacol* 2015;39:247–58. <https://doi.org/10.1016/J.ETAP.2014.11.025>.
- [3] Todeschini R and C V. *Molecular Descriptors Chemoinformatics*. vol. 1, 2009, p. 107–11. <https://doi.org/10.1002/9783527628766.fmatter1>.
- [4] Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010;29:476–88. <https://doi.org/10.1002/minf.201000061>.
- [5] Hall LH, Kier LB. Electrotological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. vol. 35. 1995.
- [6] Kier LB, Hall LH, Frazer JW. AN INDEX OF ELECTROTOPOLOGICAL STATE FOR ATOMS IN MOLECULES. vol. 7. 1991.
- [7] Todeschini R. *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing ...* - Roberto Todeschini, Viviana Consonni - Google Livros 2009. [https://books.google.com.br/books?hl=pt-BR&lr=&id=6Zp7Yrqzv8AC&oi=fnd&pg=PP2&dq=Molecular+descriptors+for+chemoinformatics,+2009.&ots=Jq3KHH\\_LJg&sig=tYodnaol9Hn8dYy0AJ9PgPbWx24&redir\\_esc=y#v=onepage&q=Molecular%20descriptors%20for%20chemoinformatics%2C%202009.&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=6Zp7Yrqzv8AC&oi=fnd&pg=PP2&dq=Molecular+descriptors+for+chemoinformatics,+2009.&ots=Jq3KHH_LJg&sig=tYodnaol9Hn8dYy0AJ9PgPbWx24&redir_esc=y#v=onepage&q=Molecular%20descriptors%20for%20chemoinformatics%2C%202009.&f=false) (accessed October 4, 2023).
- [8] Ricci-Lopez J, Aguila SA, Gilson MK, Brizuela CA. Improving Structure-Based Virtual Screening with Ensemble Docking and Machine Learning. *J Chem Inf Model* 2021;61:5362–76. [https://doi.org/10.1021/ACS.JCIM.1C00511/ASSET/IMAGES/LARGE/CI1C00511\\_0006.JPEG](https://doi.org/10.1021/ACS.JCIM.1C00511/ASSET/IMAGES/LARGE/CI1C00511_0006.JPEG).
- [9] Rücker C, Rücker G, Meringer M. Y-randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 2007;47:2345–57. <https://doi.org/10.1021/CI700157B/ASSET/IMAGES/LARGE/CI700157BF0002.JPEG>.
- [10] Yu X. Prediction of Depuration Rate Constants for Polychlorinated Biphenyl Congeners. *ACS Omega* 2019;4:15615–20. [https://doi.org/10.1021/ACSOMEGA.9B02072/ASSET/IMAGES/MEDIUM/AO9B02072\\_M010.GIF](https://doi.org/10.1021/ACSOMEGA.9B02072/ASSET/IMAGES/MEDIUM/AO9B02072_M010.GIF).

- [11] Ciura K, Ulenberg S, Kapica H, Kawczak P, Belka M, Bączek T. Drug affinity to human serum albumin prediction by retention of cetyltrimethylammonium bromide pseudostationary phase in micellar electrokinetic chromatography and chemically advanced template search descriptors. *J Pharm Biomed Anal* 2020;188:113423. <https://doi.org/10.1016/J.JPBA.2020.113423>.
- [12] Dragos H, Gilles M, Alexandre V. Predicting the predictability: A unified approach to the applicability domain problem of qsar models. *J Chem Inf Model* 2009;49:1762–76. [https://doi.org/10.1021/CI9000579/SUPPL\\_FILE/CI9000579\\_SI\\_001.ZIP](https://doi.org/10.1021/CI9000579/SUPPL_FILE/CI9000579_SI_001.ZIP).
- [13] Randic M. *Molecular Shape Profiles*1. vol. 35. 1995.
- [14] Grisoni F, Merk D, Byrne R, Schneider G. Scaffold-Hopping from Synthetic Drugs by Holistic Molecular Representation. *Sci Rep* 2018;8. <https://doi.org/10.1038/s41598-018-34677-0>.
- [15] Grisoni F, Merk D, Consonni V, Hiss JA, Tagliabue SG, Todeschini R, et al. Scaffold hopping from natural products to synthetic mimetics by holistic molecular similarity. *Communications Chemistry* 2018 1:1 2018;1:1–9. <https://doi.org/10.1038/s42004-018-0043-x>.