

Estudo de Modelos QSAR de uma Série de Inibidores da Desidroquinato Desidratase II de *Helicobacter pylori*

QSAR Models Study of a Series of Dehydroquinate Dehydratase II Inhibitors of *Helicobacter pylori*

Paulo H. de S. Miranda,^a Pedro I. C. de Oliveira,^b Estela M. G. Lourenço,^{a,c} Alexander M. S. Morais,^a Priscilla S. de S. N. Silverio,^b Euzébio G. Barbosa,^{a,b} Alessandro K. Jordão^{a,*}

^a Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Ciências Farmacêuticas, Natal-RN, Brasil

^b Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Bioinformática, Natal-RN, Brasil

^c Universidade Federal de Minas Gerais, Laboratório de Modelagem Molecular e Planejamento de Fármacos, Belo Horizonte-MG, Brasil

*E-mail: alessandro.jordao@ufrn.br

Submissão: 6 de Fevereiro de 2025

Aceite: 27 de Maio de 2025

Publicado online: 6 de Junho de 2025

The gastrointestinal disturbances caused by *Helicobacter pylori* (*H. pylori*) are responsible for a significant number of fatal cases and its prevalence and incidence are alarming worldwide. In addition, the resistance to the traditional treatments is increasing due to the development of different bacterial defence mechanisms. Therefore, studies aiming the development of new drugs are necessary. Dehydroquinatase II (DHQase II) is an enzyme that represents a promising target for the design of new antibiotic compounds. In this work, we aimed to identify the structure-activity relationship (SAR) using Activity Cliff and create different linear QSAR models based on derivatives of quinic and 3-dehydroquinic acid. A total of 56 *H. Pylori* DHQase II inhibitors were used to create three linear QSAR models, using multiple linear regression (MLR). The results of our study can guide future synthetic studies by predicting the potential antimicrobial activity before the experimental process.

Keywords: *H. pylori*; antibiotics; molecular modeling; QSAR, activity cliff.

1. Introdução

Helicobacter pylori é uma bactéria gram-negativa em forma de espiral. Ela é transmissível diretamente de pessoa para pessoa.¹ Essa bactéria é responsável por agravar muitas patologias do trato gastrointestinal, como gastrite, duodenite e úlceras gástricas.² Estima-se que cerca de metade da população mundial esteja infectada por esse microrganismo.³ O primeiro e principal tratamento para combater a infecção por *H. pylori* baseia-se em um inibidor da bomba de prótons (omeprazol) associado a dois antibióticos, claritromicina e amoxicilina.⁴ Existem outras combinações de antibióticos como opções de tratamento.⁵ No entanto, o desenvolvimento de resistência bacteriana aos antibióticos, especialmente à claritromicina, representa um obstáculo no tratamento de *H. pylori*, já que alguns pacientes podem não apresentar a melhora clínica desejada.⁵⁻⁸

A via do chiquimato consiste em sete reações catalisadas por diferentes enzimas. Essa via metabólica é responsável pela síntese de alguns aminoácidos aromáticos, folatos e vitaminas K e E.⁶ Esses nutrientes são essenciais para o desenvolvimento de bactérias como a *H. pylori*.⁷ Como essa via não existe em mamíferos, suas enzimas são consideradas alvos promissores para o planejamento racional de novos medicamentos altamente potentes e seletivos. Particularmente, a reação enzimática na via do chiquimato é um processo de desidratação catalisado pela enzima 3-desidroquinato desidratase (DHQase) e resulta na formação de 3-desidrochiquimato.⁸

Existem duas enzimas diferentes de 3-desidroquinato desidratase, DHQase tipo I ou tipo II.⁹ Embora ambas as enzimas atuem nessa desidratação reversível, elas possuem mecanismos distintos.¹⁰ A DHQase I está presente em plantas, fungos e algumas espécies bacterianas, como *Escherichia coli*, *Staphylococcus aureus* e *Salmonella typhi*. Essa enzima utiliza um mecanismo que envolve a formação de uma base de Schiff. Por outro lado, a DHQase II não envolve a formação de base de Schiff e é essencial em espécies bacterianas como *H. pylori*, que possuem apenas enzimas do tipo II e as utilizam em sua rota biossintética.^{11,12} Isso torna a enzima um alvo seletivo e atraente para o desenvolvimento de novos medicamentos para o tratamento de *H. pylori*.

No entanto, o processo de descoberta e desenvolvimento de novos medicamentos é caro, complexo e demorado. Para reduzir essas dificuldades, a indústria farmacêutica e grupos de pesquisa começaram a utilizar metodologias direcionadas.¹³ Entre essas, as aplicações de

métodos *in silico* têm aumentado nas últimas décadas e se tornaram uma ferramenta muito importante. Esses métodos são considerados racionais, permitindo a redução de custos e tempo experimental.¹⁴

A relação quantitativa estrutura-atividade (*Quantitative Structure-Activity Relationship* - QSAR) baseia-se nas características químicas (descritores moleculares) de compostos com diferentes potências para a mesma atividade biológica, a fim de construir modelos de regressão. Os modelos construídos podem prever a atividade biológica de novos compostos.¹⁵ Consequentemente, este método é eficaz para encontrar relações confiáveis entre a estrutura química e a atividade biológica, por meio dos descritores obtidos de uma série de compostos já estudados ou de novos compostos químicos. Outro método computacional importante para estudar as propriedades das moléculas é o *Activity Cliff* (AC). Nesse método, uma série de moléculas com alta similaridade estrutural são comparadas em 2D ou 3D para criar uma relação entre estrutura e atividade. É um ponto focal da exploração de SAR (*Structure-Activity Relationship*) e indica sua descontinuidade. Após a coleta de dados, podem ser propostas modificações químicas para manter alta potência e melhorar propriedades, como farmacocinética, estabilidade molecular e outras.^{16,17}

O objetivo deste estudo foi construir modelos QSAR 2D, 3D e híbridos, utilizando descritores de campo, como descritores clássicos 2D, moleculares 3D e uma combinação deles, extraídos de uma série de inibidores de DHQase II de *H. pylori*. Os respectivos modelos foram construídos usando métodos de regressão linear múltipla passo a passo (*stepwise multiple linear regression* - S-MLR), nos quais

modelos de regressão são projetados com boa capacidade preditiva, utilizando poucos descritores. Os modelos obtidos foram amplamente validados e interpretados com a análise auxiliar de AC. O índice de paisagem estrutura-atividade (*Structure Activity Landscape Index* - SALI) simplificou a interpretação das características estruturais que governam a inibição da DHQase II de *H. pylori*. Três modelos QSAR são propostos aqui, sendo que aquele com melhor capacidade preditiva e bons parâmetros estatísticos pode ajudar no planejamento racional de novos inibidores da DHQase II com boa atividade biológica.

2. Metodologia

2.1. Geração do banco de dados

Por meio de uma busca na literatura por inibidores da enzima desidroquinato desidratase II de *H. pylori*, foram encontrados 56 compostos (Tabela 1).¹⁸⁻²³ Cada modelo molecular dos compostos foi construído usando o software Marvin Sketch, versão 18.16.²⁴ A protonação molecular dos compostos foi estimada em pH = 7,40 considerando a principal microespécie. Cada molécula foi cuidadosamente revisada por meio de uma avaliação espacial utilizando o software UCSF Chimera,²⁵ seguindo o protocolo preconizado por Fourches *et al.*²⁶ Todos os modelos moleculares foram otimizados utilizando o nível de teoria semiempírica PM7 no MOPAC2016,²⁷ considerando o efeito do solvente água por meio do modelo COSMO²⁸ de solvatação implícita, utilizando uma constante dielétrica

Tabela 1. Inibidores da desidroquinato desidratase II de *H. pylori*

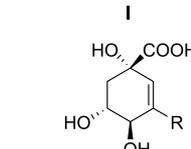
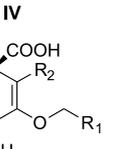
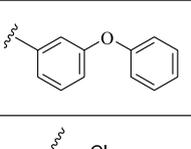
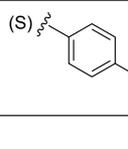
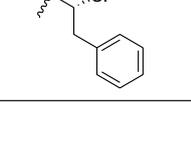
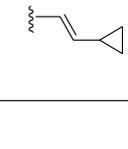
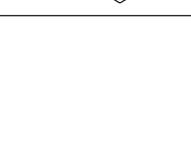
Código	R	pKi	Código	R	PKi
I A		5,92	II I	(S) 	6,01
I B		6,31	II J	(S) 	5,85
I C		4,39	II K	(S) 	6,05
I D		4,98	III A		6,22

Tabela 1. Inibidores da desidroquinato desidratase II de *H. pylori* (cont.)

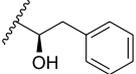
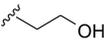
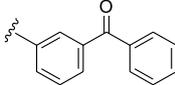
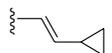
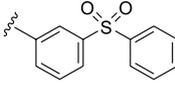
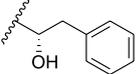
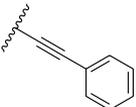
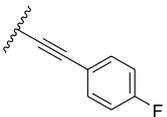
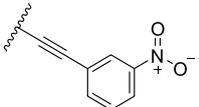
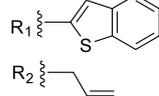
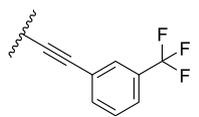
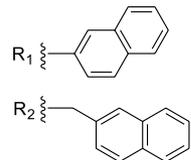
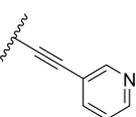
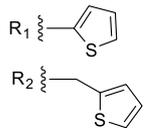
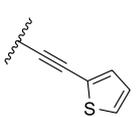
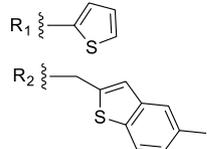
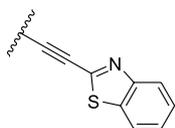
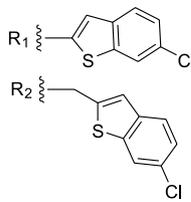
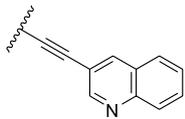
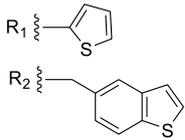
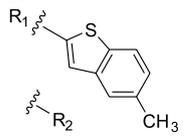
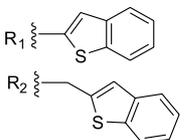
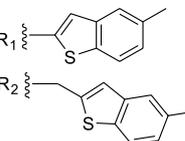
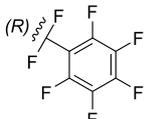
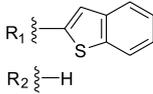
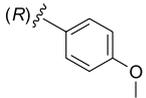
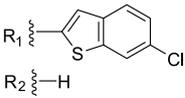
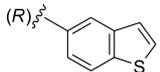
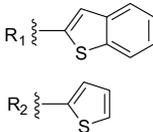
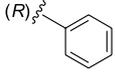
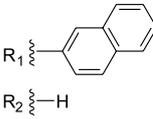
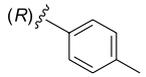
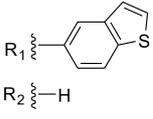
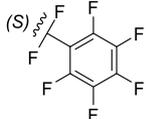
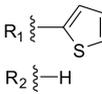
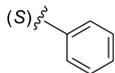
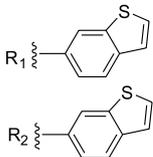
Código	R	pKi	Código	R	PKi
I E		4,43	III B	H	4,47
I F		3,55	III C	CH ₃	5,8
I G		3,73	III D		5,9
I H		5,54	III E		5,97
I I		4,62	III F		4,42
I J		4,48	III G		5,47
I K		6,6	III H		6,46
I L		6,97	III I		6,49
I M		7,01	IV A		5,96
I N		6,99	IV B		6,59
I O		6,15	IV C		4,84
I P		6,5	IV D		6,55
I Q		7,03	IV E		5,17

Tabela 1. Inibidores da desidroquinato desidratase II de *H. pylori* (cont.)

Código	R	pKi	Código	R	PKi
I R		5,92	IV F		7,0
I S		4,7	IV G		6,89
I T		5,03	IV H		7,01
II A		4,67	IV I		7,3
II B		6,01	IV J		6,88
II C		6,77	IV K		6,69
II D		6,8	IV L		6,55
II E		5,85	IV M		6,51
II F		6,6	IV N		6,78
II G		5,59	IV O		6,04
II H		5,93	IV P		6,85

As moléculas dos conjuntos de treinamento e teste estão ilustradas no material suplementar.

(ϵ) de 78,4 para representar a água a 298 K. O processo de otimização foi considerado convergido quando o gradiente RMS atingiu um valor inferior a $0,01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ e deslocamento atômico máximo de $0,01 \text{ \AA}$. Os valores de K_i em μM foram convertidos em $\text{p}K_i$ e utilizados para análises subsequentes.

2.2. Procedimentos de alinhamento

Antes da construção do modelo 3D-QSAR, foi realizado um alinhamento dos compostos utilizando o *pharmACOPhore*,²⁹ automatizado por um *shell script* ad hoc (Figura 1). Todos os 56 compostos foram sobrepostos no sítio ativo (PDB ID: 2WKS).³⁰ Para alinhar os ligantes ao sítio ativo, foi realizado um *docking* manual, uma vez que o alinhamento automático falhou em produzir conformações ligadas razoáveis. Este procedimento foi revisado para manter a qualidade do *docking* e dos modelos que seriam construídos com esses dados.

2.3. Cálculo de descritores 2D e 3D

O modelo 2D-QSAR foi construído com base nos descritores moleculares 2D calculados utilizando o software PaDEL-Descriptor,³¹ onde inicialmente foram gerados todos os descritores 2D disponíveis no software. Os descritores resultantes foram submetidos a um processo de filtragem usando o R^{32} no R-Studio versão 3.4.4³³ para remover descritores com valores ausentes e aqueles com variância nula. O pacote Caret³⁴ foi utilizado para remover descritores altamente intercorrelacionados (0,95).

Para construir o modelo 3D-QSAR, foram calculados os descritores Lennard-Jones (LJ), Eletrostático (QQ), Ligações de Hidrogênio (HB) e Hidrofóbico (HF) de acordo com o método descrito na literatura.³⁵ Os descritores do campo de interação molecular foram calculados em uma grade com resolução de 1 \AA , criando uma matriz com um

número imenso de colunas. Para simplificar os dados e possibilitar a modelagem, pontos de dados muito distantes das moléculas alinhadas foram removidos aplicando um corte de variância de 0,02. Além disso, descritores intercorrelacionados foram eliminados (0,98), mantendo-se apenas aqueles mais bem correlacionados com a variável dependente. Diferentes níveis de corte de correlação foram empregados para os descritores 2D e 3D, a fim de manter um número adequado de descritores (com redução máxima de 60%) sem comprometer a qualidade dos modelos construídos.

2.4. Construção e validação dos modelos lineares

Para reduzir a dimensionalidade dos dados das variáveis independentes, uma série de processos de filtragem foi aplicada usando a seleção ordenada de preditores (*Ordered Predictor Selection* - OPS), disponível no software QSAR Modeling.³⁶ O método de regressão linear múltipla passo a passo (*Stepwise Multiple Linear Regression* - S-MLR) implementado no software NanoBRIDGES³⁷ foi utilizado para construir os três modelos de regressão linear (2D-QSAR, 3D-QSAR e QSAR Híbrido) a partir de um conjunto de variáveis independentes reduzidas pelo OPS em cada uma das matrizes. O conjunto de treinamento, contendo 40 compostos, foi usado para construir os três modelos lineares, enquanto o conjunto de teste, com 17 compostos, foi utilizado para avaliar a capacidade preditiva dos modelos construídos (Q_{ext}^2 , Equação 1).

A qualidade dos modelos construídos foi analisada usando os parâmetros estatísticos R^2 (*o coeficiente de determinação, que mede o ajuste do modelo aos dados de treinamento*) e Q^2 (*o coeficiente de determinação da validação cruzada, que estima a capacidade preditiva interna do modelo*) (Eq. 1). A validação interna dos modelos foi realizada usando *Leave-N-out* (LNO) e *y-randomization*. O LNO foi usado para investigar a

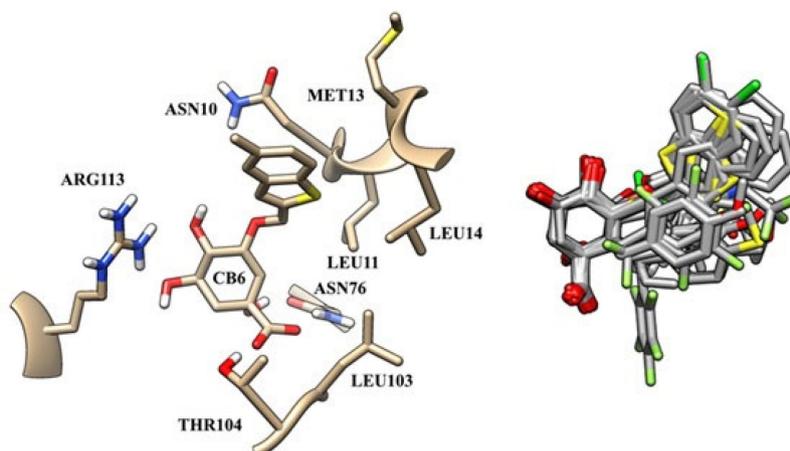


Figura 1. Representação tridimensional da proteína derivada da estrutura cristalográfica do PDB 2WKS, evidenciando o sítio de interação com o ligante (CB6). As cadeias laterais dos resíduos fundamentais para o reconhecimento molecular – incluindo Asn10, Met13, Leu11, Leu14, Arg113, Asn76, Thr104 e Leu103. À direita, observa-se o alinhamento molecular dos 56 compostos

robustez dos modelos, enquanto o *y-randomization* verificou a presença de correlações aleatórias nos modelos construídos. Adicionalmente, o Índice de Perturbação do Modelo no Domínio de Aplicabilidade (*Applicability Domain - Model Disturbance Index - AD-MDI*)³⁸ permitiu avaliar a confiabilidade preditiva dos modelos QSAR construídos.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Coeficiente de Determinação (R}^2\text{)} \quad (1a)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Coeficiente de Determinação da Validação Cruzada (Q}^2\text{) (Leave-one-out - LOO)} \quad (1b)$$

onde y_i = valor real da atividade biológica para a molécula i . \hat{y}_i é valor predito pelo modelo QSAR, \bar{y} é a média dos valores reais de atividade. n = número de moléculas no conjunto de treinamento. $\hat{y}_{(i)}$ é o valor predito para a molécula i quando ela é excluída do treinamento (método Leave-one-out). Q^2_{ext} é calculado de forma similar ao R^2 , mas as amostras do conjunto separado para validação externa.

2.5. Activity cliffs

Para investigar a notável similaridade estrutural entre os 56 compostos, o software Openbabel³⁹ foi utilizado para calcular a similaridade estrutural 2D das moléculas e fornecer os coeficientes de Tanimoto. Os resultados de similaridade desses coeficientes foram usados para calcular o Índice de Paisagem Estrutura-Atividade (*Structure Activity Landscape Index - SALI*),⁴⁰ conforme a equação 2.

$$SALI = \frac{|A_i - A_j|}{1.05 - SIM_{ij}} \quad (2)$$

A_i na equação representa a atividade biológica do composto A , enquanto A_j refere-se à atividade biológica de outro composto, e SIM_{ij} representa a similaridade estrutural entre as moléculas. O valor 1,05 na fórmula resulta de uma

modificação da equação para evitar divisões por zero. Todos os dados foram plotados no programa Gephi.⁴¹ Os valores de SALI foram atribuídos como arestas e filtrados para a interpretação dos resultados do activity cliff.

3. Resultados e Discussão

3.1. Modelo de QSAR

O método MLR é satisfatório ao lidar com um pequeno número de descritores moleculares e tem sido frequentemente destacado em estudos envolvendo QSAR por apresentar um bom poder preditivo.^{42,43} Aplicando esse método, foram criados três modelos lineares. Os modelos demonstraram um desempenho aceitável na predição da atividade biológica, embora o modelo 2D-QSAR não tenha apresentado resultados tão bons em comparação com os outros dois modelos construídos (3D-QSAR e QSAR Híbrido) (Figura 2). Todos os três modelos matemáticos estão dentro dos parâmetros estatísticos, de acordo com dados da literatura disponível, onde os valores de Q^2 e R^2 devem ser maiores que 0,5 e 0,6, respectivamente, para um modelo QSAR satisfatório (Tabela 2).⁴⁴

O Leave-N-Out (LNO) é uma das opções de validação interna frequentemente utilizada em estudos de QSAR, pois pode avaliar de forma eficaz a robustez do modelo obtido. Esse método consiste em excluir uma amostra por vez do conjunto de treinamento, criar um modelo em cada iteração sem as amostras removidas e prever a atividade biológica.⁴⁵ Seguindo esse conceito, podemos observar na Figura 3 pequenas flutuações no valor de Q^2_{LNO} quando até 15 amostras são retiradas durante o processo de validação cruzada do modelo 2D-QSAR; o modelo 3D-QSAR resistiu até a remoção de 7 amostras, enquanto o modelo QSAR Híbrido sustentou a remoção de 19 amostras.

A Y-randomization, por sua vez, tem como objetivo detectar uma possível correlação entre as variáveis dependentes e independentes, que são, respectivamente, a atividade biológica e os descritores moleculares.⁴⁶ Os valores observados de R^2 e Q^2 nesse parâmetro devem

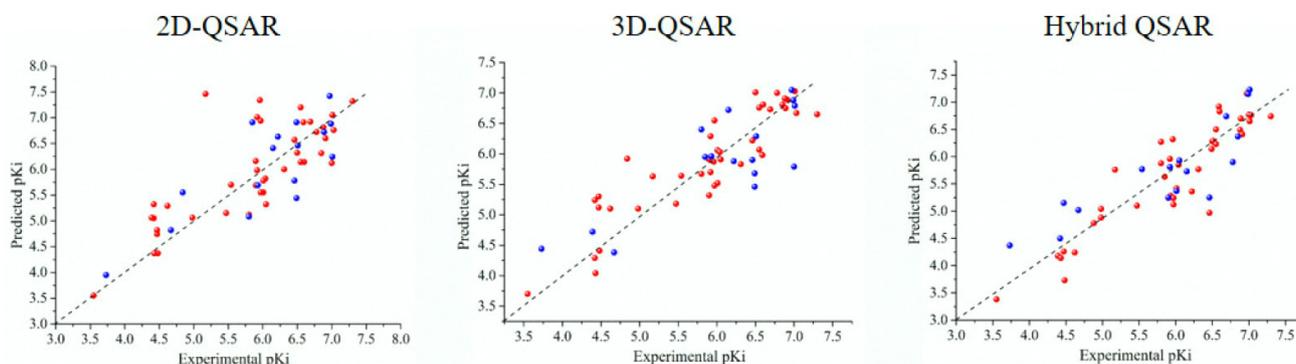
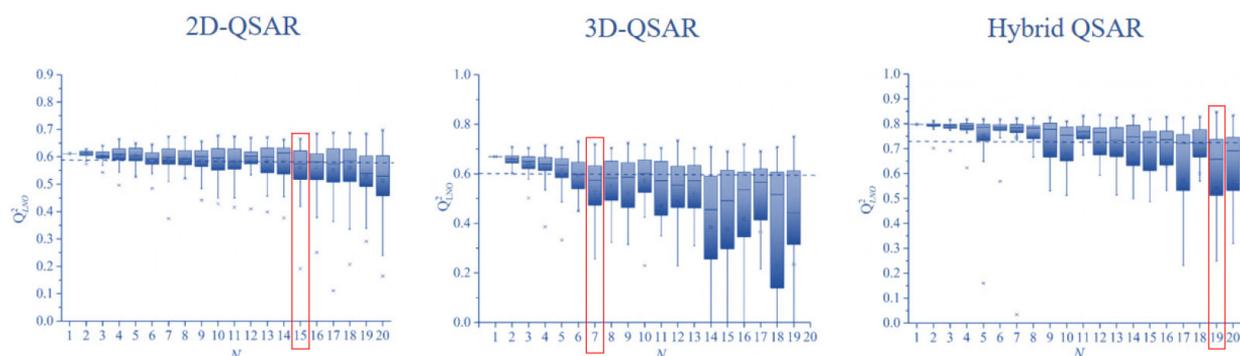


Figura 2. Modelos QSAR construídos pelo método S-MLR. Os valores previstos dos compostos inibidores são apresentados no eixo Y, enquanto os valores experimentais de pKi dos inibidores de DHQase II são apresentados no eixo X. O conjunto de treinamento está em vermelho e o conjunto de teste em azul

Tabela 2. Métricas dos respectivos modelos lineares

Parâmetros Estatísticos	R ²	Q ²	Q ² _{exp}	Equação
2D-QSAR	0,64	0,64	0,63	1,21*(topoRadius) - 0,27*(nHBint5)- 0,32*(nBondsM) +5,66*(GATS4c)+ 1,37*(piPC5)-9,98
3D-QSAR	0,78	0,76	0,74	0,05752*(HF1) -1,10869*(QQ1)- 0,07837*(LJ1)+6,90*(HF2)+0,42*(HB1)-0,81*(LJ2)- 0,017*(HB2)+0,019*(HF3) +7,09
Hybrid QSAR	0,87	0,78	0,78	0,02*(HF+)-[1,31*(HF-)-0,55*(HF-)] +0,0064(AATS8i)+4,31*(QQ) -0,015*(LJ)-1,20*(nHBDon) -0,533*(nsssCH)+10,47 (F = 22,16, p < 0,001)

**Figura 3.** A validação LNO aplicada aos respectivos modelos lineares

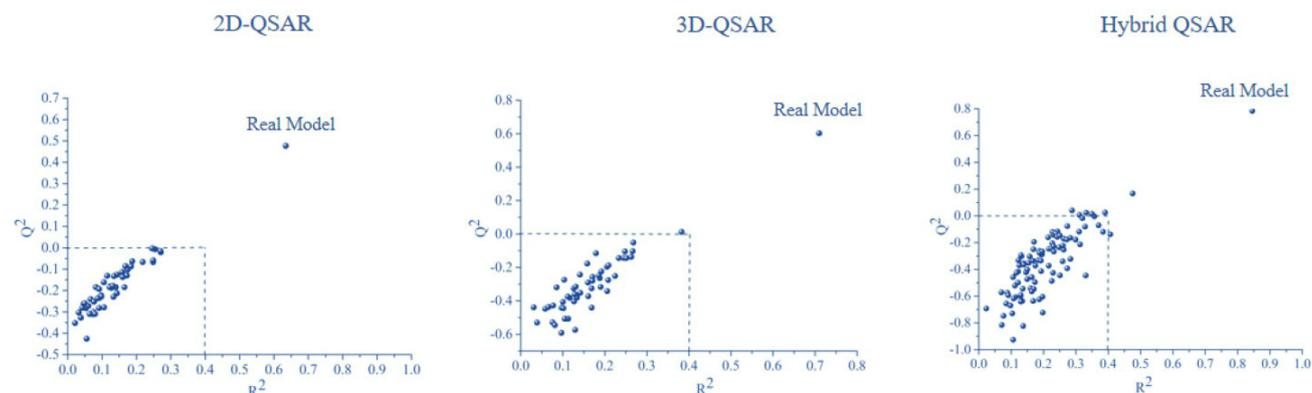
ser baixos e não apresentar nenhuma significância real, a fim de comprovar que o modelo não foi obtido ao acaso.⁴⁴ Nesse contexto, os interceptos da regressão linear (R² e Q²) apresentaram valores abaixo de 0,4 e 0,0, respectivamente, comprovando que os três modelos não foram obtidos aleatoriamente (Figura 4).

A interpretação do AD-MDI é baseada no erro de predição (PE) das amostras do conjunto de teste, utilizadas para validar a capacidade preditiva do modelo construído. Isso é combinado com a análise do Índice de Perturbação do Modelo (MDI), que se baseia nas informações dos descritores moleculares usados para construir os modelos lineares de QSAR. Para simplificar a interpretação do AD-MDI dos três modelos lineares, os resultados foram divididos em quatro quadrantes (A, B, C e D) (Figura 5).

A amostra presente no quadrante A é considerada uma

amostra não confiável, classificada como um falso positivo em termos de confiabilidade. O quadrante B é um espaço que, idealmente, deveria estar vazio, pois representa uma área com altos índices de PE e MDI, caracterizando amostras não confiáveis. Diferentemente do quadrante B, as amostras do conjunto de teste presentes no quadrante C apresentam baixos índices de PE e MDI, indicando que são confiáveis e estão dentro do domínio de aplicabilidade (AD). Por fim, o quadrante D indica amostras confiáveis, mas julgadas como não confiáveis, pois estão em uma região com baixo índice de PE, embora a mesma região apresente um alto índice de MDI. O modelo híbrido apresentou o melhor perfil de AD, com pouquíssimas amostras fora do quadrante C, que é onde a maioria das amostras deveria estar.

Assim, a compilação dos resultados de LNO e Y-randomization, juntamente com a análise do AD-MDI

**Figura 4.** A Y-randomization dos modelos preditivos. Os modelos reais são exibidos distantes dos modelos gerados com as amostras aleatórias

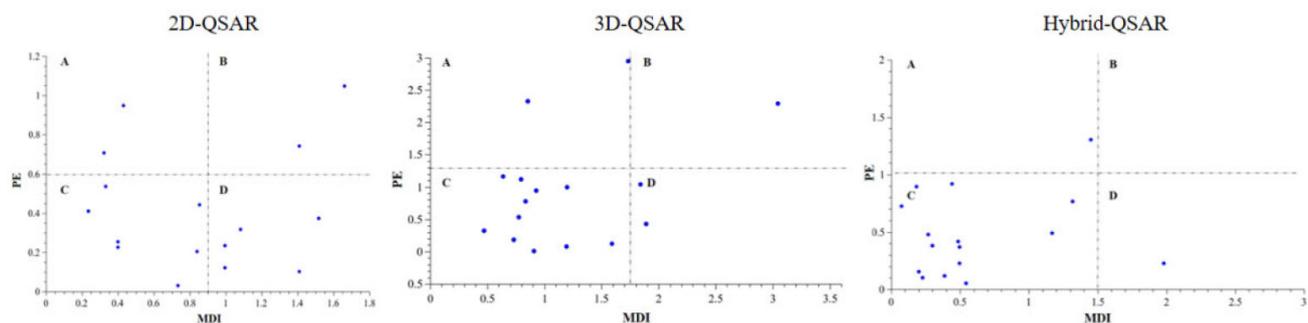


Figura 5. Estudo do domínio de aplicabilidade nos modelos lineares construídos

para os três modelos, destaca o modelo híbrido de QSAR como o mais estatisticamente significativo e robusto. Consequentemente, ele foi considerado o mais adequado para prever a atividade biológica dos compostos publicados analisados neste estudo e também foi considerado apropriado para testar novos inibidores da DHQase II. Outros estudos têm utilizado modelos mistos com resultados promissores, e como este modelo apresentou um desempenho superior aos modelos baseados em descritores de um único tipo na predição da atividade biológica das moléculas utilizadas neste trabalho, o modelo híbrido foi o único selecionado.⁴⁷

Os descritores de interação de campo molecular usados para construir o modelo final híbrido de QSAR foram representados como esferas coloridas. Cada descritor foi plotado no espaço e nomeado de acordo com sua natureza (LJ, HF e QQ) e o sinal da regressão (positivo ou negativo) de cada coeficiente de regressão (Figura 6). Um dos compostos mais ativos da série de inibidores da DHQase II ($pK_i = 7,3$) foi usado como referência para facilitar a interpretação visual dos descritores. Além dos descritores 3D, alguns descritores 2D, como AATS8i (autocorrelação média de Moreau-Broto com atraso 8 ponderada pelo potencial de ionização), nHBDon (número de doadores de ligação de hidrogênio) e nsssCH (número de átomos do tipo sssCH), também contribuíram para o desenvolvimento do modelo matemático híbrido.

Na Figura 6, os descritores de campo são representados por esferas coloridas. É importante notar que os descri-

tores HF+ (esfera verde) e HF- (esferas amarelas) não representam o mesmo descritor com sinais opostos. Eles indicam a contribuição do campo hidrofóbico (HF) em diferentes pontos específicos da grade tridimensional que foram selecionados pelo modelo S-MLR. O sinal positivo (HF+) sugere que a presença de caráter hidrofóbico naquela região espacial específica é favorável à atividade inibitória, enquanto o sinal negativo (HF-) indica que a hidrofobicidade naquela outra localização espacial é desfavorável à atividade. O sinal específico de cada descritor se encontra na Tabela 2.

3.2. Interpretação do modelo com base em evidências de activity cliffs

Os activity cliffs consistem de compostos ativos que, embora estruturalmente semelhantes, apresentam discrepâncias significativas em suas potências, evidenciando uma estreita correlação entre a estrutura química e a atividade biológica dos compostos.⁴⁸ Essas discrepâncias destacam regiões onde modificações na estrutura química se revelam vantajosas e impactantes para a atividade biológica.¹⁷ A relação estrutura-atividade (SAR) das moléculas foi interpretada de forma visual mediante a utilização do índice SALI aplicado ao diagrama de rede, no qual os respectivos valores foram atribuídos às arestas, resultando em representações gráficas. Por meio da aplicação de um filtro analítico, foram identificados três activity cliffs relevantes, conforme ilustrado nas Figuras 7 a 9.

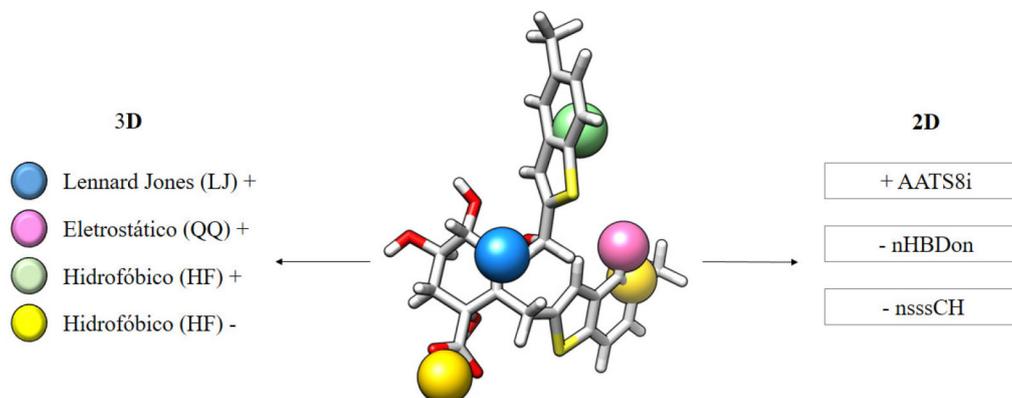


Figura 6. Descritores utilizados para construir o modelo híbrido. Os descritores 3D também são exibidos como esferas próximas à molécula. Os descritores 2D foram utilizados nos cálculos do modelo

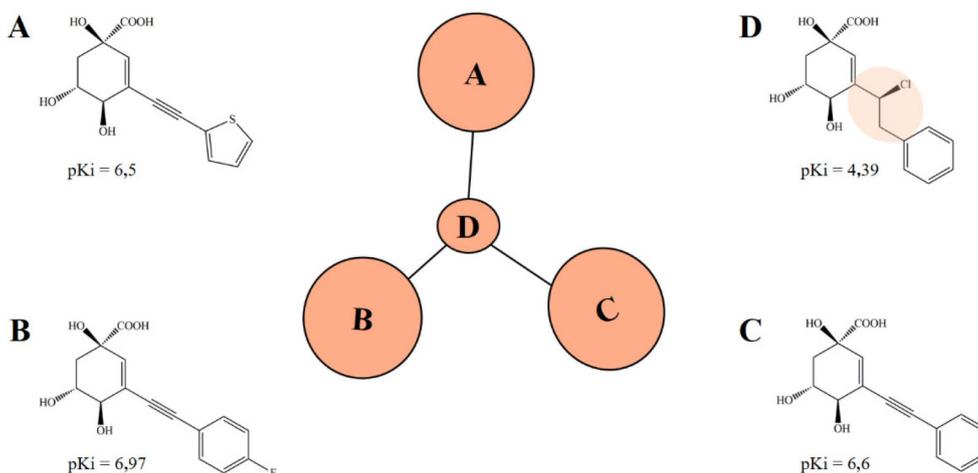


Figura 7. De acordo com a análise da relação estrutura-atividade associada às interpretações intermoleculares dos descritores, a presença da região destacada do composto D resulta em uma diminuição significativa na atividade biológica

No cliff ilustrado na Figura 7, observa-se uma diferença marcante na atividade biológica dos compostos, ainda que apresentem semelhanças estruturais. Nesse caso, a variação crucial reside no tipo de ligação química estabelecida, que exerce influência direta sobre a potência inibitória. Em particular, os inibidores da série DHQase II de *H. pylori* que possuem uma ligação tripla entre o núcleo de anidroquinato e os substituintes heterocíclicos favorecem uma interação hidrofóbica com o resíduo de aminoácido ASN 10, conforme indicado pelo descritor HF+ nessa região (Figura 6).⁴⁸ Além disso, de acordo com a química clássica, a ligação tripla é mais forte do que a ligação simples, resultando em maior estabilidade molecular durante a interação intermolecular com os resíduos de aminoácidos que circundam o sítio ativo. Esse fator contribui para potencializar a força de interação entre o composto e a sítio enzimático.

No cliff apresentado na Figura 8, evidencia-se como a variação estrutural na região hidrofóbica dos compostos interfere em seu potencial de inibição enzimática. Esse cliff é resultado de três moléculas, cuja principal diferença responsável por afetar a atividade biológica reside na extensão do radical presente em cada composto. Na molécula menos ativa, o radical não se mostra suficientemente longo para estabelecer interações hidrofóbicas com os resíduos MET 13, LEU 11 e LEU 14, componentes centrais do bolsão hidrofóbico. Em contraste, o composto mais ativo apresenta um substituinte volumoso com propriedades hidrofóbicas nessa região do sítio ativo, favorecendo o contato com os referidos resíduos. Dessa forma, a esfera azul (LJ positivo, Figura 6) destacada a relevância de grupos volumosos e hidrofóbicos nessa região.

Por fim, o cliff retratado na Figura 9 demonstra que as variações de potência inibitória entre as moléculas podem ser substancialmente atribuídas à presença de halogênios, como o cloro (Cl), evidenciado nos anéis do composto I. A nuvem eletrônica desse átomo de cloro interfere na interação do composto com resíduos próximos no bolsão hidrofóbico

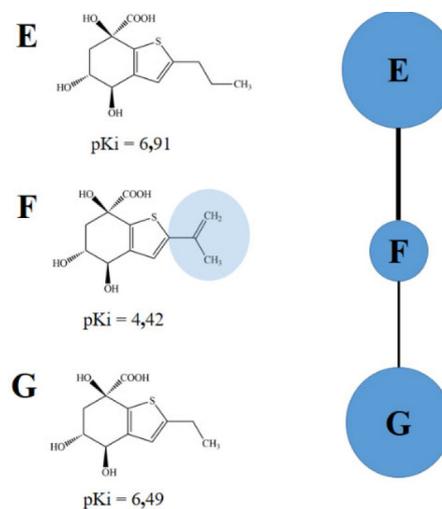


Figura 8. De acordo com a análise da relação estrutura-atividade associada às interpretações intermoleculares dos descritores, a região destacada do composto F resulta em uma diminuição significativa na atividade biológica

da proteína. Nesse contexto, o descritor QQ+ identifica uma região de interação eletrostática associada à presença de grupos químicos com cargas negativas, que interagem preferencialmente com a porção hidrofóbica do sítio ativo, resultando em uma diminuição da atividade inibitória do composto. Grupos volumosos e hidrofóbicos, sobretudo anéis aromáticos neutros, são recomendados para ocupar essa região específica, o que se condiz com a presença do descritor HF negativo nessa área.

O outro descritor HF negativo ilustrado em amarelo próximo a carboxila do núcleo de anidroquinato (Figura 6), relaciona-se aos resíduos adjacentes ao qual o ligante interage com a região formada pelos resíduos THR 104, LEU 103 e ASN 76 (Figura 1). No que se refere aos descritores bidimensionais, o AATS8i ajuda a discriminar o nível de ionização dos ligantes, enquanto nHBDn e nssCH permitem distinguir a presença de esqueletos alifáticos.

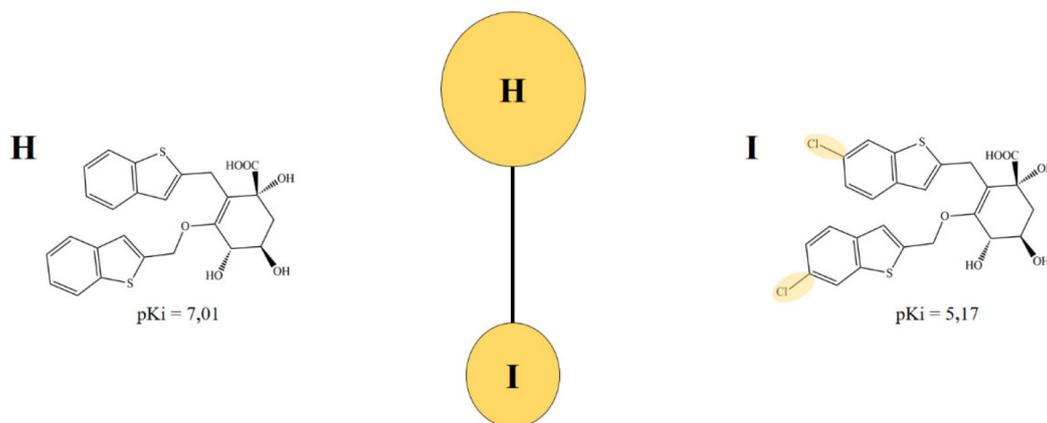


Figura 9. De acordo com a análise da relação estrutura-atividade associada às interpretações intermoleculares dos descritores, a região destacada do composto I resulta em uma diminuição significativa na atividade biológica

4. Conclusão

Este trabalho demonstrou a viabilidade e a robustez da aplicação de modelos QSAR lineares, especialmente do modelo híbrido, para a predição da atividade biológica de inibidores da DHQase II de *Helicobacter pylori*. A integração de descritores 2D e 3D permitiu capturar de forma mais abrangente as características estruturais determinantes para a atividade inibitória, resultando em um modelo estatisticamente viável e com boa capacidade preditiva. A análise de activity cliffs forneceu insights adicionais sobre modificações estruturais, contribuindo para a compreensão da relação estrutura-atividade desses compostos. Os resultados obtidos oferecem um ferramental para o planejamento racional e a triagem virtual de novos inibidores da DHQase II, potencialmente acelerando o desenvolvimento de terapias contra infecções por *H. pylori*.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Código de Financiamento 001) e ao Núcleo de Processamento de Alto Desempenho da UFRN (NPAD/IMD/UFRN) pelo apoio oferecido neste estudo.

Informações Suplementares

Esse trabalho apresenta informações suplementares e estão disponíveis gratuitamente em: <https://rvq.sbgq.org.br>

Referências Bibliográficas

1. Krzyżek, P.; Biernat, M. M.; Gościński, G.; Intensive formation of coccoid forms as a feature strongly associated with highly pathogenic *Helicobacter pylori* strains. *Folia Microbiologica* **2019**, *64*, 273. [Crossref] [PubMed]
2. Kamogawa-Schifter, Y.; Yamaoka, Y.; Uchida, T.; Beer, A.; Tribl, B.; Schöniger-Hekele, M.; Trauner, M.; Dolak, W.; Prevalence of *Helicobacter pylori* and its CagA subtypes in gastric cancer and duodenal ulcer at an Austrian tertiary referral center over 25 years. *PLoS One* **2018**, *13*, e0197695. [Crossref] [PubMed]
3. Aitila, P.; Mutyaba, M.; Okeny, S.; Kasule, M. N.; Kasule, R.; Ssedyabane, F.; Okongo, B.; Apecu, R. O.; Muwanguzi, E.; Oyet, C.; Prevalence and Risk Factors of *Helicobacter pylori* Infection among Children Aged 1 to 15 Years at Holy Innocents Children's Hospital, Mbarara, South Western Uganda. *Journal of Tropical Medicine* **2019**, *2019*, 9303072. [Crossref] [PubMed]
4. Yamaoka, Y.; Ansari, S.; Current understanding and management of *Helicobacter pylori* infection: An updated appraisal. *F1000Research* **2018**, *7*, 721. [Crossref] [PubMed]
5. Chey, W. D.; Leontiadis, G. I.; Howden, C. W.; Moss, S. F. ACG Clinical Guideline: Treatment of *Helicobacter pylori* Infection. *The American Journal of Gastroenterology* **2017**, *112*, 212. [Crossref] [PubMed]
6. Cui, D.; Deng, A.; Bai, H.; Yang, Z.; Liang, Y.; Liu, Z.; Qiu, Q.; Wang, L.; Liu, S.; Zhang, Y.; Shi, Y.; Qi, J.; Wen, T.; Molecular basis for feedback inhibition of tyrosine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. *Journal of Structural Biology* **2019**, *206*, 322. [Crossref] [PubMed]
7. Aversch, N. J. H.; Krömer, J. O.; Metabolic engineering of the shikimate pathway for production of aromatics and derived compounds—Present and future strain construction strategies. *Frontiers in Bioengineering and Biotechnology* **2018**, *6*, 32. [Crossref] [PubMed]
8. Liu, C.; Liu, Y.-M.; Sun, Q.-L.; Jiang, C.-Y.; Liu, S.-J.; Unraveling the kinetic diversity of microbial 3-dehydroquinate dehydratases of shikimate pathway. *AMB Express* **2015**, *5*, 7. [Crossref] [PubMed]
9. Tohge, T.; Watanabe, M.; Hoefgen, R.; Fernie, A. R.; Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in Plant Science* **2013**, *4*, 62. [Crossref] [PubMed]
10. Gonzalez-Bello, C.; Inhibition of Shikimate Kinase and Type II Dehydroquinase for Antibiotic Discovery: Structure-Based Design and Simulation Studies. *Current Topics in Medicinal Chemistry* **2016**, *16*, 960. [Crossref] [PubMed]

11. Lence, E.; Van der Kamp, M. W.; González-Bello, C.; Mulholland, A. J.; QM/MM simulations identify the determinants of catalytic activity differences between type II dehydroquinase enzymes. *Organic & Biomolecular Chemistry* **2018**, *16*, 4443. [[Crossref](#)] [[PubMed](#)]
12. Gourley, D. G.; Shrive, A. K.; Polikarpov, I.; Krell, T.; Coggins, J. R.; Hawkins, A. R.; Issacs, N. W.; Sawyer, L.; The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nature Structural & Molecular Biology* **1999**, *6*, 521. [[Crossref](#)] [[PubMed](#)]
13. Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *British Journal of Pharmacology* **2011**, *162*, 1239. [[Crossref](#)] [[PubMed](#)]
14. Lourenço, E. M. G.; Fernandes, J. M.; Carvalho, V. F.; Grougnet, R.; Martins, M. A.; Jordão, A. K.; Zucolotto, S. M.; Barbosa, E. G.; Identification of a selective PDE4B inhibitor from *Bryophyllum pinnatum* by target fishing study and *in vitro* evaluation of quercetin 3-O- α -L-Arabinopyranosyl-(1 \rightarrow 2)-O- α -L-Rhamnopyranoside. *Frontiers in Pharmacology* **2020**, *10*, 1582. [[Crossref](#)] [[PubMed](#)]
15. Miranda, P. H. S.; Lourenço, E. M. G.; Morais, A. M. S.; de Oliveira, P. I. C.; Silverio, P. S. S. N.; Jordão, A. K.; Barbosa, E. G.; Molecular modeling of a series of dehydroquinase dehydratase type II inhibitors of *Mycobacterium tuberculosis* and design of new binders. *Molecular Diversity* **2021**, *25*, 1. [[Crossref](#)] [[PubMed](#)]
16. Bajorath, J.; Representation and identification of activity cliffs. *Expert Opinion on Drug Discovery* **2017**, *12*, 879. [[Crossref](#)] [[PubMed](#)]
17. Stumpfe, D.; Hu, Y.; Dimova, D.; Bajorath, J.; Recent progress in understanding activity cliffs and their utility in medicinal chemistry. *Journal of Medicinal Chemistry* **2014**, *57*, 18. [[Crossref](#)] [[PubMed](#)]
18. Paz, S.; Tizón, L.; Otero, J. M.; Llamas-Saiz, A. L.; Fox, G. C.; van Raaij, M. J.; Lamb, H.; Hawkins, A. R.; Laphorn, A. J.; Castedo, L.; González-Bello, C.; Tetrahydrobenzothiophene Derivatives: Conformationally Restricted Inhibitors of Type II Dehydroquinase. *ChemMedChem* **2011**, *6*, 266. [[Crossref](#)] [[PubMed](#)]
19. Tran, A. T.; Cergol, K. M.; West, N. P.; Randall, E. J.; Britton, W. J.; Bokhari, S. A. I.; Ibrahim, M.; Laphorn, A. J.; Payne, R. J.; Synthesis and Evaluation of Potent Ene-yne Inhibitors of Type II Dehydroquinases as Tuberculosis Drug Leads. *ChemMedChem* **2011**, *6*, 262. [[Crossref](#)] [[PubMed](#)]
20. Prazeres, V. F. V.; Castedo, L.; Lamb, H.; Hawkins, A. R.; González-Bello, C.; 2-Substituted-3-dehydroquinic acids as potent competitive inhibitors of type II dehydroquinase. *ChemMedChem* **2009**, *4*, 1980. [[Crossref](#)] [[PubMed](#)]
21. Tizón, L.; Otero, J. M.; Prazeres, V. F. V.; Llamas-Saiz, A. L.; Fox, G. C.; van Raaij, M. J.; Lamb, H.; Hawkins, A. R.; Ainsa, J. A.; Castedo, L.; González-Bello, C.; A prodrug approach for improving antituberculosis activity of potent *Mycobacterium tuberculosis* type II dehydroquinase inhibitors. *Journal of Medicinal Chemistry* **2011**, *54*, 6063. [[Crossref](#)] [[PubMed](#)]
22. Payne, R. J.; Riboldi-Tunnicliffe, A.; Kerbarh, O.; Abell, A. D.; Laphorn, A. J.; Abell, C.; Design, synthesis, and structural studies on potent biaryl inhibitors of type II dehydroquinases. *ChemMedChem* **2007**, *2*, 1010. [[Crossref](#)] [[PubMed](#)]
23. Blanco, B.; Sedes, A.; Peón, A.; Otero, J. M.; van Raaij, M. J.; Thompson, P.; Hawkins, A. R.; González-Bello, C.; Exploring the water-binding pocket of the type II dehydroquinase enzyme in the structure-based design of inhibitors. *Journal of Medicinal Chemistry* **2014**, *57*, 3494. [[Crossref](#)] [[PubMed](#)]
24. ChemAxon-Ltd M (2013) ChemAxon-Ltd, MarvinSketch [[Link](#)]
25. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E.; UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605. [[Crossref](#)] [[PubMed](#)]
26. Fourches, D.; Muratov, E.; Tropsha, A.; Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *Journal of Chemical Information and Modeling* **2010**, *50*, 1189. [[Crossref](#)] [[PubMed](#)]
27. James J. P. Stewart (2016) MOPAC [[Link](#)]
28. Klamt, A.; Schüürmann, G.; COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions* **1993**, *2*, 799. [[Crossref](#)]
29. Korb, O.; Monecke, P.; Hessler, G.; Stützel, T.; Exner, T. E.; PharmACOPHORE: Multiple flexible ligand alignment based on ant colony optimization. *Journal of Chemical Information and Modeling* **2010**, *50*, 1669. [[Crossref](#)] [[PubMed](#)]
30. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M.; The protein data bank: A computer-based archival file for macromolecular structures. *Archives of Biochemistry and Biophysics* **1978**, *185*, 584. [[Crossref](#)] [[PubMed](#)]
31. Yap, C. W.; PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry* **2011**, *32*, 1466. [[Crossref](#)] [[PubMed](#)]
32. R Foundation for Statistical Computing. (2018) R: a Language and Environment for Statistical Computing. [[Link](#)]
33. Team RC (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [[Link](#)]
34. Kuhn, M.; Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **2008**, *28*, 1. [[Crossref](#)]
35. Patil, R. B.; Barbosa, E. G.; Sangshetti, J. N.; Sawant, S. D.; Zambre, V. P.; LQTA-R: A new 3D-QSAR methodology applied to a set of DGAT1 inhibitors. *Computational Biology and Chemistry* **2018**, *74*, 123. [[Crossref](#)] [[PubMed](#)]
36. Martins, J. P. A.; Ferreira, M. M. C.; QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. *Química Nova* **2013**, *36*, 554. [[Crossref](#)]
37. Ambure, P.; Aher, R. B.; Gajewicz, A.; Puzyn, T.; Roy, K.; "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems* **2015**, *147*, 1. [[Crossref](#)]
38. Yan, J.; Zhu, W.-W.; Kong, B.; Lu, H.-B.; Yun, Y.-H.; Huang, J.-H.; Liang, Y.-Z.; A combinational strategy of model disturbance and outlier comparison to define applicability domain in quantitative structural activity relationship. *Molecular Informatics* **2014**, *33*, 503. [[Crossref](#)] [[PubMed](#)]

39. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R.; Open Babel: An Open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 33. [[Crossref](#)] [[PubMed](#)]
40. Guha, R.; Exploring Structure-Activity Data Using the Landscape Paradigm. *Wiley Interdisciplinary Reviews Computational Molecular Science* **2012**, *2*, 829. [[Crossref](#)] [[PubMed](#)]
41. Bastian, M.; Heymann, S.; Jacomy, M.; Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference* **2009**, 3,361. [[Crossref](#)]
42. Sabet, R.; Khabnadideh, S.; Fathi, D.; Emami, L.; QSAR and Docking Study of Isatin Analogues as Cytotoxic Agents. *Journal of Pharmaceutical Research International* **2019**, *27*, 1. [[Crossref](#)]
43. Rajathej, D. M.; Parthasarathy, S.; Selvaraj, S.; QSAR Analysis of Multimodal Antidepressants Vortioxetine Analogs Using Physicochemical Descriptors and MLR Modeling. *Current Computer-Aided Drug Design* **2019**, *15*, 294. [[Crossref](#)] [[PubMed](#)]
44. Kiralj, R.; Ferreira, M. M. C.; Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *Journal of the Brazilian Chemical Society* **2009**, *20*, 770. [[Crossref](#)]
45. Baumann, D.; Baumann, K.; Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *Journal of Cheminformatics* **2014**, *6*, 47. [[Crossref](#)] [[PubMed](#)]
46. Rucker, C.; Rucker, G.; Meringer, M.; Y-Randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling* **2007**, *47*, 2345. [[Crossref](#)] [[PubMed](#)]
47. De Oliveira, P. I. C.; Miranda, P. H. S.; Lourenço, E. M. G.; Silverio, P. S. S. N.; Barbosa, E. G.; Planning new Trypanosoma cruzi CYP51 inhibitors using QSAR studies. *Molecular Diversity* **2021**, *25*, 2219. [[Crossref](#)] [[PubMed](#)]
48. Dimova, D.; Bajorath, J.; Advances in activity cliff research. *Molecular Informatics* **2016**, *35*, 181. [[Crossref](#)] [[PubMed](#)]