



ADRIANA CÁSSIA DE FARIA

**COMPUTER-ASSISTED IMPROVEMENT OF
SULFONYLUREAS WITH ANTIFUNGAL PROPERTIES AND
LIMITED HERBICIDAL ACTIVITY: POTENTIAL
APPLICATION IN FORAGE CONSERVATION**

**LAVRAS - MG
2022**

ADRIANA CÁSSIA DE FARIA

**COMPUTER-ASSISTED IMPROVEMENT OF SULFONYLUREAS WITH
ANTIFUNGAL PROPERTIES AND LIMITED HERBICIDAL ACTIVITY:
POTENTIAL APPLICATION IN FORAGE CONSERVATION**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração Química Computacional Aplicada na Agricultura, para a obtenção do título de Mestre.

Prof. Dr. Matheus Puggina de Freitas

Orientador

**LAVRAS - MG
2022**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Faria, Adriana Cássia de.

Computer-Assisted Improvement of Sulfonylureas with
Antifungal Properties and Limited Herbicidal Activity : Potential
Application in Forage Conservation / Adriana Cássia de Faria. -
2022.

93 p. : il.

Orientador(a): Matheus Puggina de Freitas.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2022.

Bibliografia.

1. Sulfonilureias. 2. Aceto-hidroxiácido Sintase. 3. QSAR. I.
Freitas, Matheus Puggina de. II. Título.

ADRIANA CÁSSIA DE FARIA

**MELHORAMENTO ASSISTIDO COMPUTACIONALMENTE DE
SULFONILUREIAS COM PROPRIEDADES ANTIFÚNGICAS E ATIVIDADE
HERBICIDA LIMITADA: POTENCIAL APLICAÇÃO NA CONSERVAÇÃO DE
FORRAGEM**

**COMPUTER-ASSISTED IMPROVEMENT OF SULFONYLUREAS WITH
ANTIFUNGAL PROPERTIES AND LIMITED HERBICIDAL ACTIVITY:
POTENTIAL APPLICATION IN FORAGE CONSERVATION**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração Química Computacional Aplicada na Agricultura, para a obtenção do título de Mestre.

APROVADA EM 10 de fevereiro de 2022.

Prof. Dra. Carla Luiza da Silva Ávila
Prof. Dra. Daiana Teixeira Mancini

UFLA
InSILICO

Prof. Dr. Matheus Puggina de Freitas
Orientador

**LAVRAS – MG
2022**

A Deus, pelo dom da vida; aos meus pais, Adilson e Fátima, por todo apoio e incentivo; à minha família, pelo total apoio nessa caminhada vitoriosa; ao meu noivo, Patrich, por toda paciência e amor; aos meus colegas, que sempre estão presentes.

Dedico

AGRADECIMENTOS

Agradeço primeiramente a Deus, pelo dom da vida e por me dar forças para ultrapassar todos os desafios.

Agradeço aos meus pais Adilson e Fátima e os meus irmãos Ricardo e Patrícia pelo carinho, atenção e apoio que me deram durante toda a minha vida.

Ao meu noivo Patrigh que sempre esteve comigo me apoiando, incentivando e me dando forças para a realização deste trabalho.

Agradeço ao meu orientador Prof. Dr. Matheus Puggina de Freitas pela excelente orientação, paciência, apoio, incentivo e disposição durante a realização deste trabalho.

Agradeço a professora Elaine Fontes Ferreira da Cunha por sua contribuição nos estudos de docking molecular.

Agradeço aos professores participantes da banca examinadora Prof^ª. Dra. Carla Luiza da Silva Ávila e Prof^ª. Dra. Daiana Teixeira Mancini pela disponibilidade e pelas valiosas contribuições para este trabalho.

Agradeço a todos os professores do DQI/UFLA que direta ou indiretamente contribuíram para o meu desenvolvimento durante a Pós-Graduação.

À todos os colegas do laboratório de Química Computacional por todos os ensinamentos.

À Universidade Federal de Lavras e ao Departamento de Química pela oportunidade concedida para realização do mestrado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – pela concessão da bolsa de mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro.

Muito obrigada!

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.” (Arthur Schopenhauer)

RESUMO GERAL

A conservação de forragem para alimentação animal – silagem – requer o uso de antifúngicos como aditivos para garantir alta eficiência fermentativa em ambiente anaeróbico. Nesse contexto, novos compostos pertencentes à classe das sulfonilureias foram recentemente sintetizados com o propósito de modificar as estruturas químicas de substâncias análogas e já conhecidas, melhorando suas propriedades antifúngicas e minimizando suas atividades herbicidas. O presente trabalho propõe o estudo em nível molecular dessas sulfonilureias, com o objetivo de se encontrarem os sítios quimiofóricos (porções da estrutura molecular) responsáveis pela atividade antifúngica e pela seletividade (atividade antifúngica/atividade herbicida) das moléculas a serem estudadas. Para isso, utilizou-se a metodologia computacional de modelagem molecular MIA-QSAR (*Multivariate Image Analysis applied to Quantitative Structure-Activity Relationships*), bem como a ferramenta MIA-Plot (para identificação dos sítios quimiofóricos). O método MIA-QSAR foi empregado para encontrar uma correlação entre as estruturas químicas das sulfonilureias com suas respectivas propriedades biológicas. Como resultado, os modelos MIA-QSAR foram confiáveis, robustos e preditivos, ou seja, para atividade antifúngica, a média dos principais parâmetros de validação foi $r^2 = 0,936$, $q^2 = 0,741$ e $r^2_{\text{pred}} = 0,720$, enquanto para atividade herbicida, o modelo foi preditivo ($r^2_{\text{pred}} = 0,981$ e $r^2_{\text{m}} = 0,944$). Dessa forma, e sabendo-se quais partes das estruturas químicas afetam as propriedades biológicas e de que maneira (aumentando ou diminuindo a atividade microbiológica), foi possível propor 46 novas estruturas químicas com propriedades melhoradas, das quais, 9 apresentaram índices de seletividade calculados promissores. Estudos de *docking* foram realizados com a finalidade de validar as previsões de QSAR e compreender os mecanismos de ação dos compostos em estudo.

Palavras-Chave: Sulfonilureias. Aceto-hidroxiácido Sintase. QSAR. Fungicida. Forragem.

GENERAL ABSTRACT

The conservation of forage for animal feed - silage - requires the use of antifungals as additives to ensure high fermentative efficiency in an anaerobic environment. In this context, new compounds belonging to the class of sulfonylureas have recently been synthesized with the purpose of modifying the chemical structures of analogous and already known substances, improving their antifungal properties and minimizing their herbicidal activities. This work proposes the study at the molecular level of these sulfonylureas, with the objective of finding the chemophoric sites (portions of the molecular structure) responsible for the antifungal activity and for the selectivity (antifungal activity/herbicidal activity) of the molecules to be studied. Accordingly, we employed the computational methodology of molecular modeling MIA-QSAR (Multivariate Image Analysis applied to Quantitative Structure-Activity Relationships), as well as the MIA-Plot tool (to identify the chemophoric sites). The MIA-QSAR method was used to find a correlation between the chemical structures of the sulfonylureas with their respective biological properties. As a result, the MIA-QSAR models were reliable, robust, and predictive, *i.e.* for antifungal activity, the average values for the main validation parameters were $r^2 = 0.936$, $q^2 = 0.741$, and $r^2_{\text{pred}} = 0.720$, whereas for herbicidal activity, the model was predictive ($r^2_{\text{pred}} = 0.981$ and $r^2_{\text{m}} = 0.944$). In this way, and knowing which parts of chemical structures affect the biological properties and how (increasing or decreasing microbiological activity), it was possible to propose 46 new chemical structures with improved properties, from which, 9 presented promising calculated selectivity indexes. Docking studies were carried out in order to validate the QSAR predictions and understanding the mechanisms of action of the compounds under study.

Keywords: Sulfonylureas. Acetohydroxyacid synthase. QSAR. Fungicide. Forage.

LISTA DE ILUSTRAÇÕES

PRIMEIRA PARTE

Figura 1 - Estrutura química básica das sulfonilureias utilizadas neste trabalho.	17
Figura 2 - Estrutura química do herbicida chlorsulfuron.	25
Figura 3 - Chlorsulfuron, uma sulfonilureia, complexado à enzima aceto-hidroxiácido sintase (AHAS) de uma levedura e de uma planta	26
Figura 4 - Estrutura química de uma sulfonilureia contendo uma porção alquenila (destacada em vermelho) supostamente responsável pela seletividade antifúngica.	27
Figura 5 - Sistemas simulados. (a) Sistema 1: apenas a enzima AHAS. (b) Sistema 2: AHAS e cofatores FAD, Mg ²⁺ e K ⁺ . (c) Sistema 3: AHAS e cofatores ThDP, FAD, Mg ²⁺ e K ⁺	28
Figura 6 - Mecanismo de ação: Inibição irreversível da AHAS.....	29
Figura 7 - Estruturas químicas de inibidores comerciais de AHAS.....	30
Figura 8 - (a) Sobreposição de imagens utilizadas no modelo MIA-QSAR original, baseado em wireframes; (b) sobreposição de imagens utilizadas no modelo aug-MIA-QSAR.	36
Figura 9 - MIA-Plots baseados nos parâmetros b e VIP para moléculas anti-HCV.	37
Figura 10 - Imagens sobrepostas das 23 sulfonilureias estudadas neste trabalho.	51
Figura 11 - Fluxo de trabalho MIA-QSAR: conversão de imagem em uma matriz de descritores e regressão PLS com y.	51
Figura 12 - Estrutura cristalina de aceto-hidroxiácido sintase de levedura em complexada com um herbicida de sulfonilureia, sulfometuron metil	54
Figura 13 - Estrutura cristalina de <i>B. napus</i> AHAS (<i>Bn</i> AHAS) obtida por modelagem por homologia	54

SEGUNDA PARTE

ARTIGO

Figure 1 - Chemical scaffold of the sulfonilureas studied herein. Changes in R and X affect the antifungal and herbicidal activities of these agrochemicals	64
Figure 2 - MIA-QSAR workflow from image conversion to a descriptors matrix and PLS regression with y.....	65

Figure 3 - Superimposed images of the 23 sulfonylureas studied in this work. Note that the coincident substructure of the compounds is congruently aligned.....	68
Figure 4 - Correlation plots of experimental \times calculated antifungal and herbicidal activities for the 23 sulfonylurea derivatives using MIA descriptors based on r_{vdw} , ϵ , and r_{vdw}/ϵ	71
Figure 5 - MIA-plots indicating how and how much the sulfonylurea substituents affect the antifungal and herbicidal activities.....	74
Figure 6 - Ramachandran plot demonstrating the quality of the homology modeling for <i>BnAHAS</i>	78
Figure 7 - Important ligand-enzyme interactions obtained from the docking studies	80
Figure S1 - Applicability domain tests	86
Figure S2 - MIA-plots obtained from MIA descriptors encoding r_{vdw} , ϵ , and r_{vdw}/ϵ , and related to antifungal (pEC_{50}) and herbicidal (<i>I</i>) activities.	91
Figure S3 - Sequence alignment used in the homology modeling. Target sequence: <i>BnAHAS</i> . Template sequences: 3EA4. The line with “*” indicates regions of similarity between the two sequences (89.0% identity).....	92
Figure S4 - Docking score energies for the compound data set and proposals.....	93

LISTA DE TABELAS

PRIMEIRA PARTE

Tabela 1 - Exemplos de herbicidas comerciais da classe de sulfonilureias	23
Tabela 2 - Análogos de sulfonilureias e suas respectivas atividades antifúngicas.....	47
Tabela 3 - Análogos de sulfonilureias e suas respectivas atividades herbicidas.....	48
Tabela 4 - Série de 23 sulfonilureias com atividade antifúngica contra <i>C. cornigerum</i> (pEC ₅₀), atividade herbicida contra <i>B. napus</i> (<i>I</i>) e índice de seletividade (SI).....	50
Tabela 5 - Valores de pixel originais do programa GaussView e os respectivos valores proporcionais a propriedades atômicas pelos quais foram substituídos.....	52

SEGUNDA PARTE

ARTIGO

Table 1 - The series of 23 sulfonylureas with antifungal activity against <i>C. cornigerum</i> (pEC ₅₀), herbicidal activity against <i>B. napus</i> (<i>I</i>), and selectivity index (SI).....	67
Table 2 - Statistical results of the MIA-QSAR models obtained using the four test set compounds of the literature (4 , 9 , 10 , and 14).....	72
Table 3 - Statistical results (mean of 20 cycles) of the MIA-QSAR models obtained through a bootstrapping procedure with four randomly selected test set samples.	73
Table 4 - Proposed sulfonylurea candidates based on the MIA-plot outcomes and the combination of substructures of the most suitable compounds previously synthesized. ^a The calculated data correspond to mean values obtained from the three MIA-QSAR models, ranked <i>i</i> in descending order of selectivity.....	75
Table S1 - Experimental and predicted pEC ₅₀ for the MIA-QSAR models without test set. ..	87
Table S2 - Experimental and predicted pEC ₅₀ for the MIA-QSAR models with test set.....	88
Table S3 - Experimental and predicted <i>I</i> for the MIA-QSAR models without test set.....	89
Table S4 - Experimental and predicted <i>I</i> for the MIA-QSAR models with test set.....	90

LISTA DE SIGLAS E ABREVIATURAS

AHAS	Aceto-hidroxiácido Sintase
aug-MIA-QSAR	<i>Augmented Multivariate Image Analysis applied to Quantitative Structure-Activity Relationship</i>
aug-MIA-QSPR	<i>Augmented Multivariate Image Analysis applied to Quantitative Structure-Property Relationship</i>
A.S.	<i>A. solanum</i>
BCAA	Biossíntese de aminoácidos de cadeia ramificada
Bmp	<i>Bitmap 24bits</i>
<i>Bn</i> AHAS	<i>Brassica napus</i> AHAS
c_r^2 _p	Coefficiente de determinação de penalização ajustado
C.A.	<i>C. arachidicola</i>
C.C.	<i>C. cornigerum</i>
C.L.	<i>C. lagenarium</i>
CCC	<i>Concordance Correlation Coefficient</i>
FAD	Dinucleotídeo de Flavina Adenina
<i>I</i>	% de inibição na escala decimal
IMIs	Imidazolinonas
K ⁺	Íon Potássio
LOOCV	<i>leave-one-out cross-validation</i>
LV	<i>Latent Variable</i>
MAE	<i>Mean Absolute Error</i>
MIA-QSAR	<i>Multivariate Image Analysis applied to Quantitative Structure-Activity Relationships</i>
MIA-QSPR	<i>Multivariate Image Analysis applied to Quantitative Structure-Property Relationship</i>
MIA-Plot	Gráficos MIA
Mg ²⁺	Íon Magnésio
MLR	<i>Multiple Linear Regression</i>
MVD	<i>Molegro Virtual Docker</i>
PDB	<i>Protein Data Bank</i>

pEC ₅₀	Concentração necessária de um composto para que a metade da resposta máxima seja obtida
PLS	<i>Partial Least Squares</i>
P.P.	<i>P. piricola</i>
PTBs	Pirimidiniltiobenzoatos
QSAR	<i>Quantitative Structure-Activity Relationship</i>
Q ² _{F1} , Q ² _{F2} e Q ² _{F3}	<i>Predictive Squared Correlation Coefficients</i>
q ²	Coefficiente de determinação da validação cruzada
r ²	Coefficiente de determinação
r ² _{cal}	Coefficiente de determinação da calibração
r ² _{ext}	Coefficiente de determinação da validação externa
r ² _m	Coefficiente de determinação modificado
RGB	<i>Red, Green and Blue</i>
RMSE	<i>Root Mean Square Error</i>
ScAHAS	<i>Saccharomyces cerevisiae</i> aceto-hidroxiácido sintase
SCTs	Sulfanilamida-carbonil-tiazolidinonas
SI	Índice de seletividade
S.S.	<i>S. sclerotium</i>
SUs	Sulfonilureias
ThDP	Difosfato de tiamina
TPs	Triazolopirimidina-sulfonamidas
VIP	<i>Variable Importance in Projection</i>
2D	Bidimensional
3D	Tridimensional

LISTA DE SÍMBOLOS

%	Porcentagem
ϵ	Eletronegatividade
r_{vdW}	Raio de van der Waals
r_{vdW}/ϵ	Razão entre o raio de van der Waals e a eletronegatividade
Σ	Somatório
\leq	Menor ou igual que
$>$	Maior que

SUMÁRIO

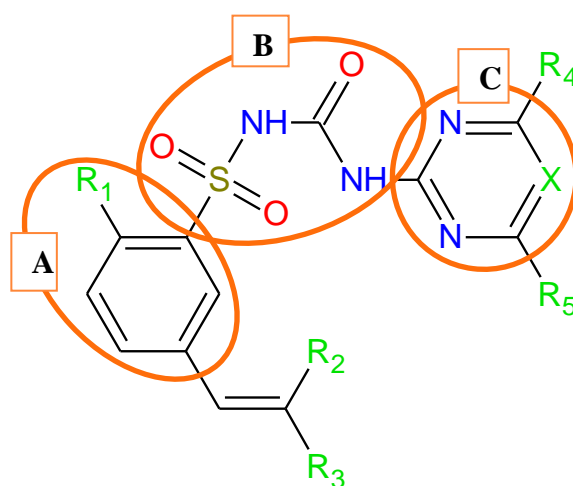
PRIMEIRA PARTE	16
1 INTRODUÇÃO GERAL	17
2 OBJETIVOS	20
2.1 Objetivo Geral	20
2.2 Objetivos Específicos	20
3 REFERENCIAL TEÓRICO	21
3.1 Silagem	21
3.2 Herbicidas	22
3.3 Sulfonilureias como agentes antifúngicos e herbicidas	23
3.4 AHAS	27
3.5 Relação quantitativa entre estrutura química e atividade biológica – QSAR	31
3.6 MIA-QSAR	34
3.7 Validação de modelos QSAR	38
3.7.1 Coeficiente de Correlação de Concordância	41
3.7.2 Erro Médio Absoluto	42
3.7.3 k e k'	43
3.7.4 Q^2_{F1} , Q^2_{F2} e Q^2_{F3}	44
3.8 Docking molecular	45
4 METODOLOGIA	47
5 CONSIDERAÇÕES GERAIS	56
REFERÊNCIAS	57
SEGUNDA PARTE	62
ARTIGO - Computer-assisted improvement of sulfonilureas with antifungal properties and limited herbicidal activity: Potential application in forage conservation	63
1. INTRODUCTION	64
2. MATERIAL AND METHODS	66
3. RESULTS AND DISCUSSION	70
ACKNOWLEDGMENTS	81
REFERENCES	81
APPENDIX A – SUPPORTING INFORMATION	85

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

As sulfonilureias (FIGURA 1) compreendem uma classe de substâncias químicas utilizadas na agricultura, particularmente como herbicidas, em decorrência da sua capacidade de interferir na biossíntese de certos aminoácidos das plantas (DUGGLEBY; MCCOURT; GUDDAT, 2008). Em geral, são compostos por um anel benzênico *orto*-substituído (A), ponte de sulfonilureia (B) e um anel heterocíclico (C) (CHEN et al., 2019; LIU, Y. C. et al., 2016).

Figura 1 — Estrutura química básica das sulfonilureias utilizadas neste trabalho.



Fonte: Do autor, 2021.

Os compostos à base de sulfonilureias também têm sido explorados como agentes antifúngicos, sendo alvos muito atraentes para o projeto de novos herbicidas, agentes antimicrobianos e agentes antifúngicos, em razão de sua elevada atividade inibitória da enzima aceto-hidroxiácido sintase (AHAS) - (EC 2.2.1.6), (LEE et al., 2013; WEI et al., 2016a).

Após uma série de modificações estruturais, descobriu-se que as sulfonilureias possuem atividade antifúngica, em razão de sua afinidade relativamente alta pela enzima AHAS. Algumas das estruturas reportadas na literatura apresentam atividade com taxa moderada/boa contra fungos patogênicos de plantas (CHEN et al., 2015; LIU, Y. C. et al., 2016; WEI et al., 2016b). De acordo com Chen et al. (2015), em uma dosagem de 100 mg L⁻¹, a maioria dos compostos alvo apresentaram atividades moderadas a favoráveis na inibição do crescimento micelial de todos os fungos de teste, entre eles, *Rhizoctonia solanii*, *Pseudomonas syringae*, *Phytophthora capsici*, *Corynespora cassiicola*, *Botrytis cinerea* e

Fusarium oxysporum; por sua vez, o herbicida de sulfonilureia comercializado, Nicosulfuron, quase não mostrou atividade fungicida.

Foi relatado que diferentes substituintes no anel pirimidina podem resultar em importantes agentes antifúngicos, até mesmo de forma inesperada (CHEN et al., 2015). A propriedade antifúngica confere potencial às sulfonilureias para sua utilização como aditivos antifúngicos no processo de silagem; no entanto, por também apresentarem atividade herbicida, modificações estruturais devem ser feitas de forma a beneficiar a atividade antifúngica em detrimento da atividade herbicida. Essa seletividade (aumento da atividade antifúngica e diminuição do efeito herbicida) é desejável, uma vez que permite a alta eficiência fermentativa em ambiente anaeróbico durante o processo de conservação da forragem para alimentação animal, sem que se altere a integridade do material vegetal.

O presente trabalho tem como objetivo o planejamento racional de novas sulfonilureias que tenham atividade antifúngica elevada e, ao mesmo tempo, uma limitada atividade herbicida. Ainda, propôs-se a avaliação das propriedades estruturais das sulfonilureias que estejam correlacionadas à sua atividade antifúngica, atividade herbicida e seletividade (atividade antifúngica/atividade herbicida). Para tanto, modelos computacionais para previsão das referidas propriedades foram construídos com base em uma aproximação conhecida como QSAR (do inglês *Quantitative Structure-Activity Relationships*). De forma concisa, os métodos QSAR consistem em correlacionar descritores moleculares obtidos experimentalmente ou teoricamente com as respectivas propriedades de uma série de moléculas (no presente caso, sulfonilureias com atividades antifúngica e herbicida conhecidas). O modelo matemático então obtido foi utilizado para prever as propriedades de novas moléculas, ainda não sintetizadas, bem como para auxiliar na interpretação acerca das subestruturas moleculares (partes da molécula) que exercem influência na atividade biológica das sulfonilureias. Usualmente, o tratamento dos dados é feito utilizando-se estatística multivariada. O método empregado neste trabalho é conhecido como MIA-QSAR (MIA = *Multivariate Image Analysis*), cujos descritores moleculares derivam dos pixels que formam as imagens das estruturas químicas das sulfonilureias. A série de sulfonilureias que compõem o modelo MIA-QSAR, bem como suas respectivas atividades biológicas, estão disponíveis na literatura (WEI et al., 2016a).

Uma vez que a resistência desenvolvida pelos microorganismos aos antifúngicos é um problema real, que pode afetar a eficiência dos aditivos à base de sulfonilureia aplicados nas plantas forrageiras, o desenvolvimento de novos princípios ativos torna-se imprescindível. O *design* racional de novas estruturas moleculares com potencial aplicação na agricultura

precede as etapas de síntese e testes físico-químico-biológicos, sendo um processo essencial para garantir a inserção de um novo produto de qualidade no mercado. A execução do presente trabalho se justifica tomando-se como base a constante necessidade de novos agentes antifúngicos, dado o possível surgimento de espécies resistentes e, também, porque é parte integrante do *workflow* das pesquisas sobre desenvolvimento de fármacos e agroquímicos.

2 OBJETIVOS

2.1 Objetivo Geral

O presente trabalho tem como objetivo o planejamento de novas sulfonilureias que apresentem atividade antifúngica elevada frente a diversos microorganismos e, ao mesmo tempo, uma limitada atividade herbicida. Para alcançar esse objetivo, utilizou-se a metodologia computacional de modelagem molecular MIA-QSAR, bem como a ferramenta MIA-Plot, para identificar a correlação entre as estruturas químicas das sulfonilureias com suas respectivas propriedades biológicas.

2.2 Objetivos Específicos

- 1) Construir um modelo MIA-QSAR com elevada capacidade preditiva da atividade antifúngica e limitada atividade herbicida para uma série de herbicidas da classe das sulfonilureias;
- 2) Interpretar o modelo com base nos coeficientes de regressão PLS dos descritores MIA, com o propósito de compreender quais regiões e substituintes na molécula mais afetam a bioatividade;
- 3) Realizar estudos de *docking* molecular, para validar os resultados e propostas advindas da análise QSAR, bem como compreender o modo de interação entre as moléculas propostas e o alvo biológico;
- 4) Divulgar os resultados por meio de publicação em revista internacional com fator de impacto relevante para a área.

3 REFERENCIAL TEÓRICO

3.1 Silagem

A alimentação de animais ruminantes é baseada em pastagens cultivadas pelo homem ou de origem nativa. Durante o período chuvoso, a alimentação disponível é abundante e apresenta boa qualidade nutricional, porém, durante o período de seca, a qualidade e a disponibilidade são reduzidas. O armazenamento da forragem como silagem é uma alternativa que tem como finalidade manter a boa qualidade e a disponibilidade durante o período de escassez da pastagem (AMORIM et al., 2019).

Silagem é uma forragem conservada obtida por meio da fermentação de forragem geralmente úmida, utilizada para alimentação animal (UEGAKI et al., 2017). A silagem deve ser produzida a partir de material vegetal com teor de umidade adequado, de aproximadamente 50% a 60%. Além disso, ela requer uma fermentação anaeróbica, que começa em aproximadamente 48 horas após o silo ter sido preenchido. A fermentação se dá, usualmente, pela inoculação com microorganismos que aumentam a velocidade de fermentação ou melhoram a silagem resultante; os inoculantes utilizados são algumas bactérias de ácido lático (a mais comum é a *Lactobacillus plantarum*). Contudo, a eficiência do processo fermentativo pode ser prejudicada pela presença de fungos, de tal forma que o uso de aditivos químicos, incluindo substâncias antifúngicas, pode auxiliar na qualidade do processo de ensilagem e no aumento da estabilidade aeróbia (KLEINSCHMIT; SCHMIDT; KUNG JUNIOR, 2005).

Um dos principais fatores que pode afetar negativamente a fermentação da silagem é a presença de oxigênio, sendo que sua manutenção necessita de condições anaeróbias estáveis (UEGAKI et al., 2017). A deterioração aeróbia de forragem para a alimentação animal resulta no aumento do número de microorganismos indesejáveis capazes de produzir substâncias tóxicas prejudiciais aos animais e aos humanos. A utilização de aditivos ajuda a melhorar significativamente a estabilidade aeróbica, a fermentação e a recuperação de matéria seca (SILVA et al., 2015).

Durante o cultivo das plantas forrageiras, os agroquímicos são normalmente aplicados para o controle de pragas, podendo contaminar e afetar a qualidade da silagem resultante. Dessa forma, o desenvolvimento de aditivos que sejam seletivos à atividade antifúngica, em detrimento da atividade herbicida, por exemplo, são especialmente desejáveis para uso no processo de fermentação da forragem para alimentação animal.

3.2 Herbicidas

Os herbicidas representam a classe de agroquímicos mais utilizada no mundo para o controle e manejo de plantas daninhas que podem afetar áreas de culturas agrícolas, áreas urbanas e outros locais; a sua ampla utilização reflete um grande avanço na agricultura (XAVIER et al., 2018).

O ciclo de cultivo é diferente entre as diversas plantas, indicando que diferentes espécies vegetais podem ser plantadas em um mesmo local e podem apresentar estágios de desenvolvimento distintos após o mesmo tempo cronológico decorrido. A variação da produção de uma dada cultura pode ser influenciada por diversos fatores, dentre eles, condições climáticas, tratamentos com herbicidas, condições do solo e as variedades de cultivo (MAYEROVÁ; MADARAS; SOUKUP, 2018).

Entre os fatores agravantes da diminuição da produtividade, as ervas daninhas são grandes responsáveis por causar impactos e perdas significativas na produtividade da cultura se não forem controladas de forma adequada (MAYEROVÁ; MADARAS; SOUKUP, 2018). Sua infestação nas culturas reduzem significativamente a produtividade e a qualidade da produção, resultando, assim, em consideráveis perdas econômicas (VAN EVERT et al., 2017).

Devido a esses fatores, novos compostos têm sido explorados e, entre eles, os compostos à base de sulfonilureias se apresentam como alvos atraentes para o planejamento de novos herbicidas, agentes antimicrobianos e agentes antifúngicos, por apresentar elevada atividade inibitória da enzima aceto-hidroxiácido sintase (AHAS) (LEE et al., 2013; WEI et al., 2016b).

Fatores como a absorção, distribuição, metabolismo e excreção (ADME) devem ser levados em consideração para o planejamento de novos compostos inibidores da aceto-hidroxiácido sintase – enzima alvo comum para tal classe de herbicidas. O mecanismo cinético do herbicida na planta é fundamental, pois o simples fato de um herbicida atingir as folhas e, ou, ser aplicado no solo não é suficiente para que ele exerça a sua ação biológica; o herbicida precisa ser absorvido, translocar-se na planta e chegar até a organela celular (alvo biológico) (WU et al., 2019).

Atualmente, as estratégias de controle de ervas daninhas tornaram-se mais dependentes de herbicidas (KUMAR et al., 2020). Os herbicidas representam a principal ferramenta utilizada no controle de ervas daninhas na agricultura moderna e são altamente eficazes contra a maioria das plantas indesejáveis. O uso massivo ou inadequado de herbicidas no controle de plantas daninhas podem causar resistência a eles, ou seja, uma forte

seleção de qualquer característica, permitindo que as populações de plantas sobrevivam e se reproduzam sob o uso recorrente de herbicidas, fazendo-se necessário o uso de herbicidas com diferentes mecanismos de ação (REY-CABALLERO et al., 2017).

Embora o uso de herbicidas desempenhe um papel fundamental na produção agrícola, é possível que, no âmbito da silagem, o herbicida danifique o material vegetal durante o processo de conservação da forragem para alimentação animal. Nesse sentido, o desenvolvimento de sulfonilureias seletivas, que possuam atividade fungicida, mas que não alterem a integridade do material vegetal, é de grande interesse econômico para o setor agropecuário.

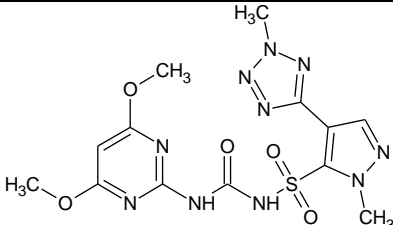
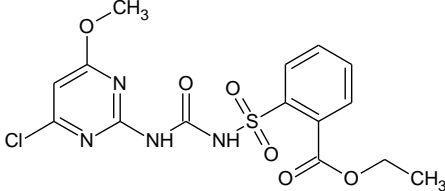
3.3 Sulfonilureias como agentes antifúngicos e herbicidas

Atualmente, a resistência de ervas daninhas aos herbicidas, especialmente a mutação de genes, tem desafiado o uso de herbicidas inibidores de AHAS comerciais. Portanto, há a necessidade do desenvolvimento de novos herbicidas (ZHOU et al., 2020).

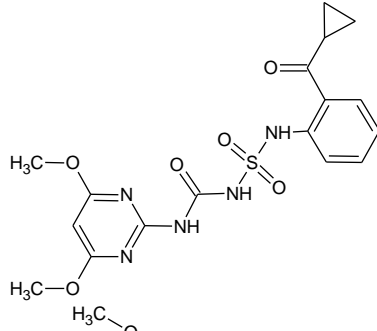
A classe de herbicidas sulfonilureias (FIGURA 1) foi desenvolvida na década de 1970 (WU et al., 2019). Sua descoberta representa um dos mais relevantes avanços da pesquisa sobre herbicidas em décadas, em razão de seus níveis de atividade herbicida sem precedentes, baixa toxicidade em mamíferos, baixa dosagem de aplicação, boa seletividade e propriedades ambientais desejáveis (HAY, 1990; WEI et al., 2017; ZHOU et al., 2020).

A Tabela 1 apresenta alguns exemplos de herbicidas comerciais pertencentes à classe das sulfonilureias.

Tabela 1 — Exemplos de herbicidas comerciais da classe de sulfonilureias (continua).

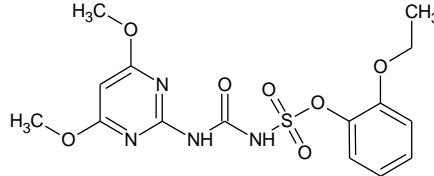
Nomes Comuns	Estrutura	Nomes Comerciais
Azimsulfuron		Gulliver
Chlorimuron-ethyl		Caput, Chlorimuron Agripec, Classic, Clorim, Clorimuron Master Nortox, Clorimuron Prentiss, Clorimuron 250 BR, Conquest, Garbor, Panzer, Smart, Staron, Stilo, Twister

Cyclosulfamuron



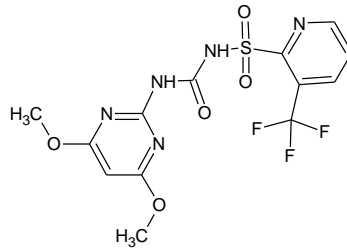
Invest

Ethoxysulfuron



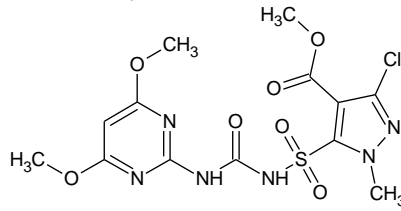
Gladium

Flazasulfuron



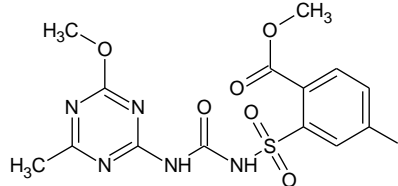
Katana

Halosulfuron-methyl



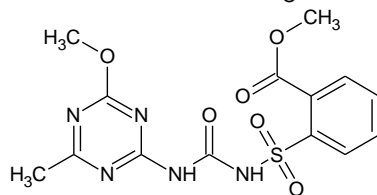
Sempra

Iodosulfuron-methyl



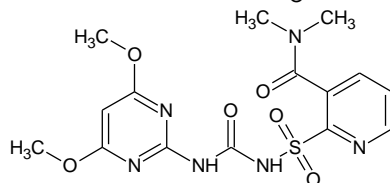
Hussar

Metsulfuron-methyl



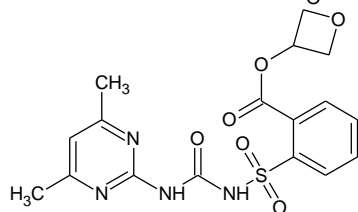
Accurate, Ally, Nufuron,
Wolf, Zartan

Nicosulfuron



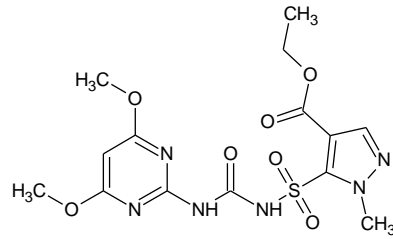
Accent, Loop, Nicosulfuron
Nortox, Nippon, Nisshin,
Pramilho, Sanson

Oxasulfuron



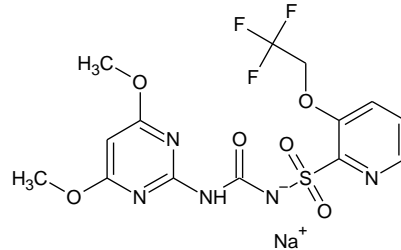
Chart

Pyrazosulfuron-ethyl



Sirius

Trifloxysulfuron-sodium



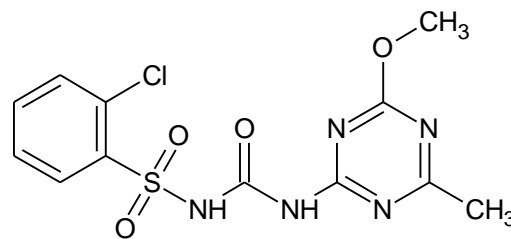
Envoke

(conclusão)

Fonte: (Adaptado de OLIVEIRA JUNIOR; CONSTANTIN; INOUE, 2011).

Um exemplo importante de herbicida representante da classe das sulfonilureias é o *chlorosulfuron* (FIGURA 2). A atividade biológica das sulfonilureias decorre da inibição da enzima AHAS, que é a enzima chave da biossíntese de aminoácidos ramificados de plantas superiores, algas, bactérias e fungos (DUGGLEBY; PANG, 2000). Portanto, além de propriedades herbicidas, espera-se que as sulfonilureias também apresentem propriedades antifúngicas.

Figura 2 — Estrutura química do herbicida *chlorsulfuron*.

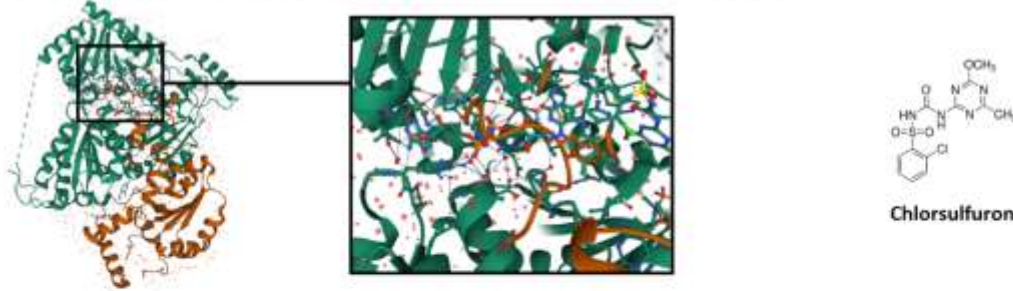


Fonte: Do autor, 2021.

A Figura 3 ilustra a interação de uma sulfonilureia – o *chlorsulfuron* – e a enzima AHAS.

Figura 3 — *Chlorsulfuron*, uma sulfonilureia, complexado à enzima aceto-hidroxiácido sintase (AHAS) de uma levedura e de uma planta.

Chlorsulfuron complexado à enzima aceto-hidroxiácido sintase de levedura - PDB: 1T9B



Chlorsulfuron complexado à enzima aceto-hidroxiácido sintase de *Arabidopsis thaliana* (planta) - PDB: 1YHZ

Fonte: Protein Data Bank (J. A. McCourt, S. S. Pang, L. W. Guddat, R. G. Duggleby, R.G., Crystal structure of yeast acetohydroxyacid synthase in complex with a sulfonilurea herbicide, chlorsulfuron, *Biochemistry*, 2005, **44**, 2330-2338; J. A. McCourt, S. S. Pang, J. King-Scott, L. W. Guddat, R. G. Duggleby, R.G., Crystal structure of *Arabidopsis thaliana* Acetohydroxyacid synthase In Complex With A Sulfonilurea Herbicide, *Proc Natl Acad Sci USA*, 2006, **103**, 569-573).

Os compostos de sulfonilureia apresentam baixa citotoxicidade e alta afinidade pela AHAS; sendo assim, as sulfonilureias são um excelente ponto de partida para o desenvolvimento de novos antifúngicos (ZHOU et al., 2017).

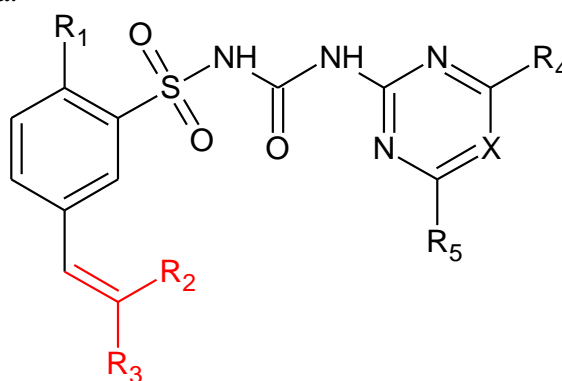
Os herbicidas a base de sulfonilureias se destacam por apresentarem alta eficiência no controle de plantas daninhas dicotiledôneas. No entanto, estudos revelam que em diferentes espécies cultivadas, a tolerância a um mesmo herbicida pode ser diferente, e essa característica pode comprometer a produtividade, causando prejuízos aos produtores (PAGNONCELLI JUNIOR et al., 2018).

A escolha de um herbicida para uma determinada cultura é realizada observando-se os níveis de tolerância dos diferentes cultivares a esse herbicida. Estudos têm demonstrado que os diferentes níveis de tolerância entre cultivares a um herbicida podem comprometer totalmente o desenvolvimento da planta. A diferença no nível de tolerância se deve à variação

genética que ocorre entre as populações; assim, enfatiza-se a importância de estudos prévios que avaliem a resposta de cultivares aos herbicidas. A seletividade aos herbicidas é um diferencial de tolerância da planta a um tratamento específico, ou seja, considera-se que quanto maior a diferença de tolerância entre a cultura e a planta daninha, maior será a segurança da aplicação (PAGNONCELLI JUNIOR et al., 2018).

Durante o processo de silagem, deseja-se que os aditivos químicos à base de sulfonilureias exibam atividade antifúngica, mas não herbicida, para não danificar o material vegetal que compõe a forragem (CHEN et al., 2015). A introdução de uma porção alquenila em uma série de sulfonilureias parece estar associada à atividade antifúngica, mas não à herbicida, sendo, portanto, seletivas a esse tipo de microorganismo (*Alternaria solani*, *Sclerotinia sclerotiorum*, *Cercospora arachidicola*, *Physalospora piricola*, *Ceratobasidium cornigerum* e *Colletotrichum lagenarium*) (WEI et al., 2016a). Sendo assim, uma série de sulfonilureias contendo uma porção alquenila (FIGURA 4) será empregada no presente trabalho para realização de estudos QSAR, com a finalidade de se modelar as atividades biológicas e correlacioná-las com as variações estruturais presentes nas moléculas.

Figura 4 — Estrutura química de uma sulfonilureia contendo uma porção alquenila (destacada em vermelho) supostamente responsável pela seletividade antifúngica.



Fonte: Do autor, 2021.

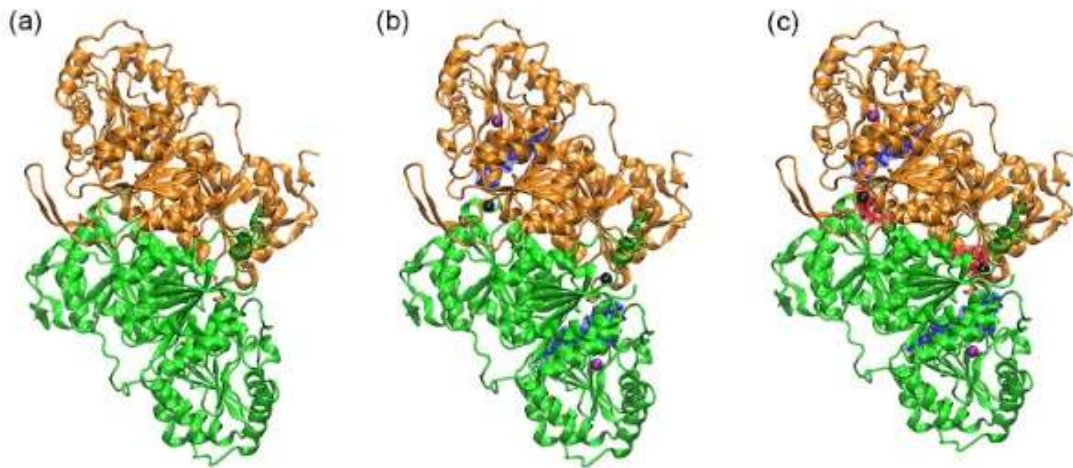
3.4 AHAS

A aceto-hidroxiácido sintase, também conhecida como acetolactato sintase – ALS (EC 2.2.1.6), é a primeira enzima na via de biossíntese de aminoácidos de cadeia ramificada (BCAA) em plantas superiores, fungos e bactérias, e não está presente em humanos e animais (LIU, Y. C. et al., 2016; WEI et al., 2016b).

A estrutura molecular dessa proteína envolve duas subunidades distintas: uma subunidade catalítica (responsável pela catálise) e uma subunidade reguladora (responsável

por modular a atividade da subunidade catalítica) (FIGURA 5). Durante sua atividade, a aceto-hidroxiácido sintase requer os seguintes cofatores: difosfato de tiamina (ThDP), dinucleotídeo de flavina adenina (FAD), um íon magnésio (Mg^{2+}) no sítio ativo e, geralmente, um íon potássio (K^+), para garantir a estabilidade estrutural (IERICH et al., 2015; LONHIENNE et al., 2016).

Figura 5 — Sistemas simulados. (a) Sistema 1: apenas a enzima AHAS. (b) Sistema 2: AHAS e cofatores FAD, Mg^{2+} e K^+ . (c) Sistema 3: AHAS e cofatores ThDP, FAD, Mg^{2+} e K^+ . As cadeias A e B são mostradas em verde e laranja, respectivamente, e os cofatores ThDP, FAD, Mg^{2+} e K^+ são mostrados em vermelho, azul, preto e roxo, respectivamente.

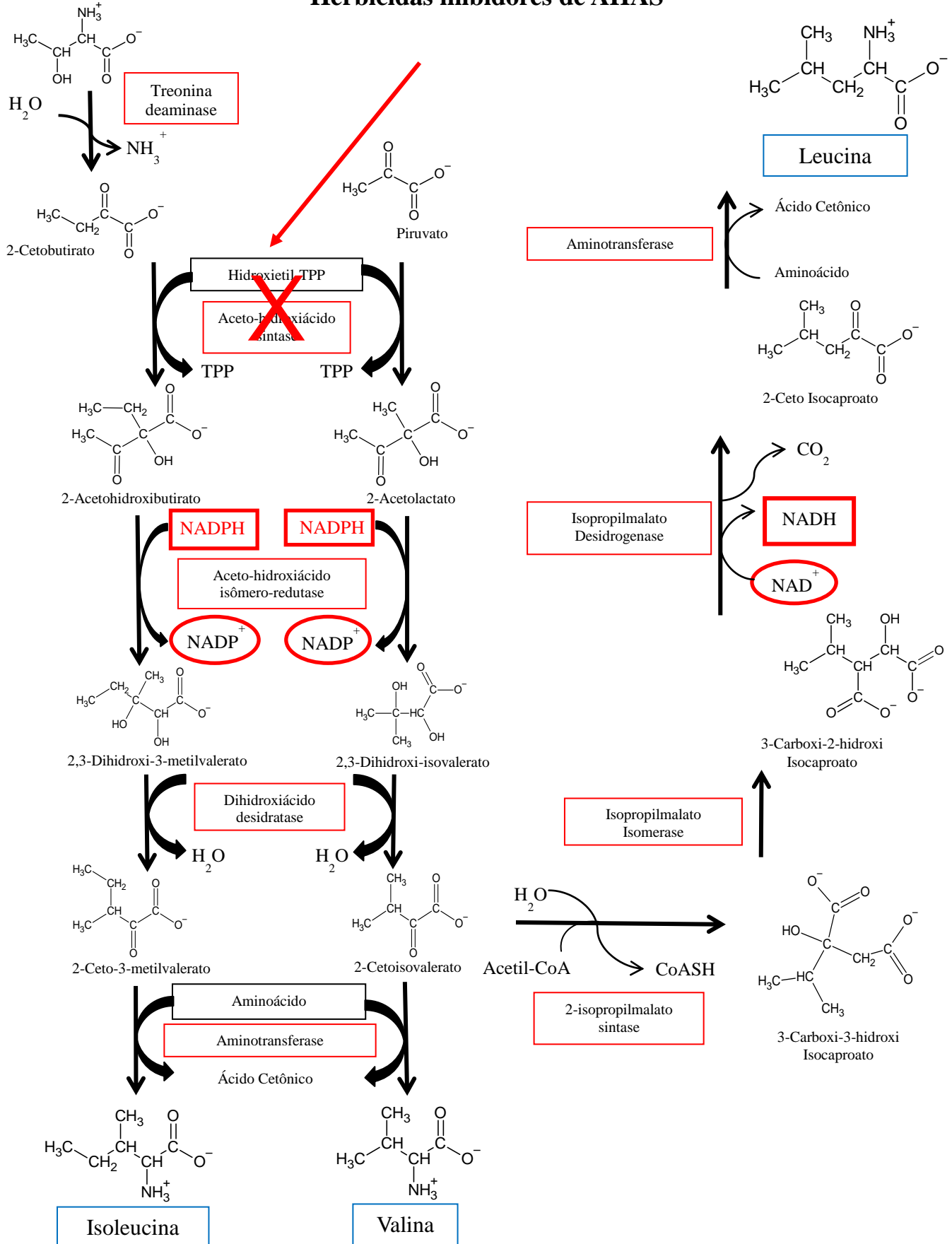


Fonte: (IERICH et al., 2015).

A AHAS é uma enzima determinante para a catálise da biossíntese de três aminoácidos essenciais nas plantas: a isoleucina, a leucina e a valina (CHEN et al., 2019). Sua inibição interrompe o ciclo catalítico e a síntese de proteínas no meristema, e bloqueia a divisão celular, resultando na inibição do crescimento interno e na morte da planta (ervas prejudiciais) (FIGURA 6) (LIU; LI; WANG, 2016; WEI et al., 2016b; ZHOU et al., 2020).

Figura 6 — Mecanismo de ação: Inibição irreversível da AHAS.

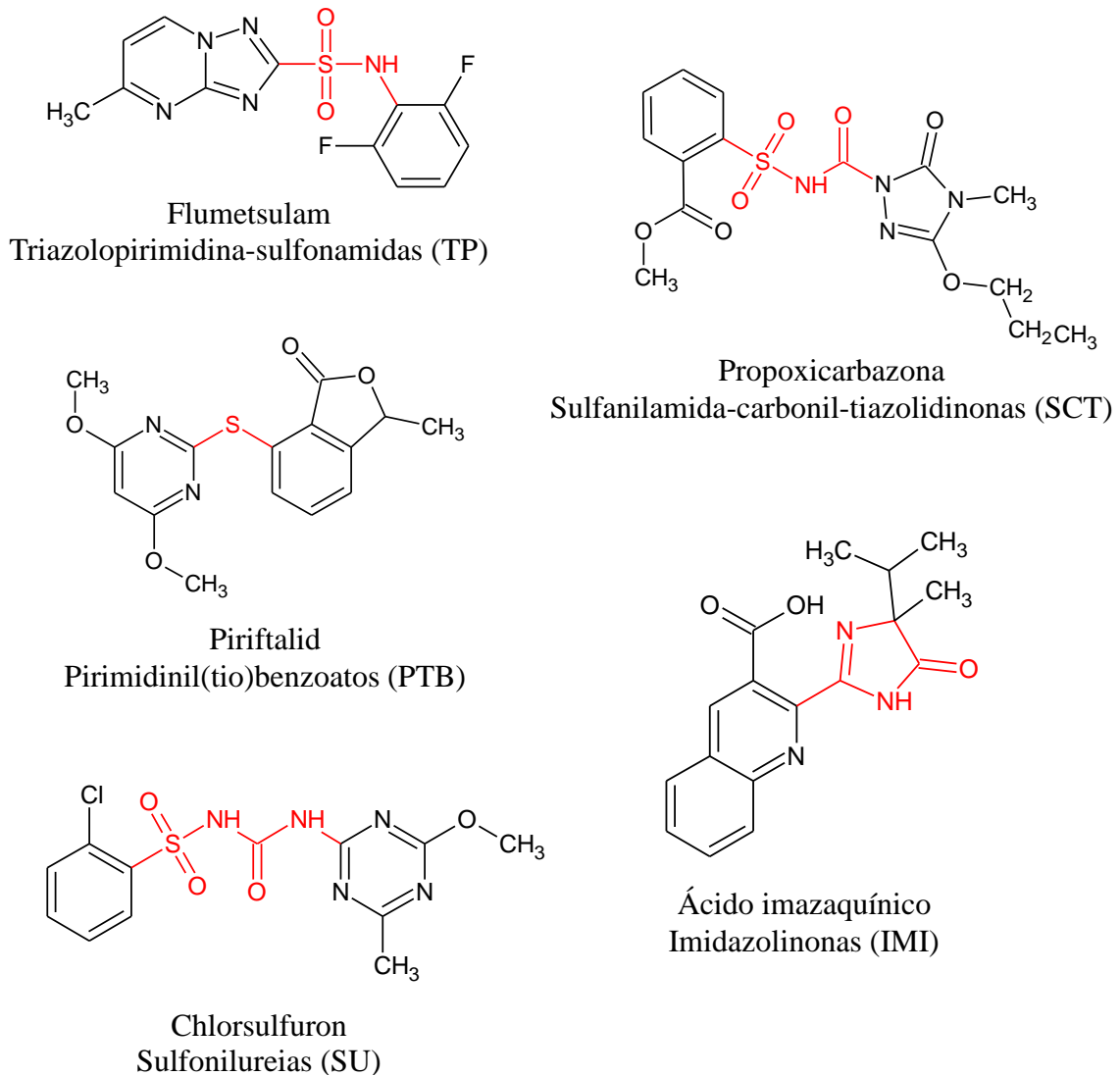
Herbicidas inibidores de AHAS



Fonte: (Adaptado de OLIVEIRA JUNIOR; CONSTANTIN; INOUE, 2011).

Dentre as diversas classes de compostos que têm a AHAS como alvo biológico, destacam-se: triazolopirimidina-sulfonamidas (TPs), sulfonilureias (SUs), pirimidinil-(tio)benzoatos (PTBs), sulfanilamida-carbonil-tiazolidinonas (SCTs) e imidazolinonas (IMIs) (FIGURA 7) (GARCIA et al., 2018; QU et al., 2017).

Figura 7 — Estruturas químicas de inibidores comerciais de AHAS.



Fonte: (QU et al., 2017).

O uso excessivo e a longo prazo de herbicidas inibidores de AHAS têm induzido o surgimento de uma elevada resistência em ervas daninhas; uma das explicações para esse recorrente evento é a existência de um único sítio de ligação na enzima AHAS. A resistência de ervas daninhas a inibidores de AHAS é causada, principalmente, por mutações de ponto

único na região do local de ligação, causando uma redução na afinidade de ligação entre o AHAS e seus inibidores (QU et al., 2017).

No entanto, a avaliação das propriedades estruturais das sulfonilureias que possam estar correlacionadas à sua atividade antifúngica, atividade herbicida e seletividade (atividade antifúngica/atividade herbicida) vão ser analisadas com base nos métodos QSAR.

3.5 Relação quantitativa entre estrutura química e atividade biológica – QSAR

Os modelos de regressão multivariada são responsáveis por relacionar quantitativamente uma variável dependente y (resposta) a um bloco de variáveis preditoras X , na forma de uma equação matemática $y = f(X)$, na qual as variáveis preditoras podem ser determinadas experimentalmente ou computacionalmente. Dentre as relações quantitativas $X \times y$ mais conhecidas, podem ser citadas as relações quantitativas estrutura-atividade (QSAR), que estão inseridas no termo mais geral “relações quantitativas estrutura-propriedade” (QSPR), em que y é uma resposta biológica e qualquer um dos preditores são designados como descritores (KIRALJ; FERREIRA, 2009).

QSAR (do inglês *Quantitative Structure-Activity Relationships*) é uma metodologia quantitativa que tem como objetivo encontrar um modelo matemático que correlacione as características estruturais químicas de compostos (codificado na forma de descritores) com uma propriedade de interesse (bioatividade) (DARÉ; RAMALHO; FREITAS, 2018; GOODE-ROMERO; AGUAYO-ORTIZ; DUEÑAS, 2019; LEELANANDA; LINDERT, 2016). Em essência, todas as abordagens QSAR implicam, direta ou indiretamente, no princípio de similaridade simples, que por muito tempo forneceu uma base para a química medicinal experimental: espera-se que compostos com estruturas semelhantes tenham atividades biológicas semelhantes (TROP SHA, 2010).

Em termos gerais, os dois objetivos principais do QSAR podem ser identificados como o desenvolvimento de uma equação ou modelo matemático e a explicação das características químicas modeladas codificadas na forma de descritores (ROY; KAR; DAS, 2015a).

A técnica QSAR foi consolidada com o trabalho pioneiro de Hansch e Fujita (1964), no qual os autores buscaram correlacionar a estrutura química de alguns ácidos benzoicos com a sua respectiva atividade pesticida. Nesse trabalho, foi estudada a correlação existente entre os valores de atividade biológicas de uma série de compostos com suas respectivas propriedades físico-químicas para a identificação e a caracterização de fatores físico-químicos que influenciavam na atividade biológica.

Nos dias atuais, a técnica QSAR é aceita como ferramenta essencial para o desenvolvimento de moléculas com diferentes propósitos, sejam elas fármacos, agroquímicos ou materiais com propriedades diferenciadas. Tal metodologia tem como essência a ideia de que para uma classe de novos compostos ou compostos não-testados, que apresentem similaridades estruturais a alguma substância já conhecida, apresentarão grandes chances de possuir atividade biológica parecida (TODESCHINI et al., 2020).

Muitos trabalhos envolvendo análises QSAR foram desenvolvidos, sendo que os mais recentes apresentam diversos métodos estatísticos sofisticados, tanto para calibração quanto para validação dos modelos construídos. Tais modelos são amplamente aplicados na avaliação dos impactos potenciais de produtos químicos, materiais, e nanomateriais na saúde humana e sistemas ecológicos. Devido a essa vasta gama de ferramentas estatísticas disponíveis, algumas publicações recentes buscam oferecer guias sobre boas práticas em QSAR (TROPISHA, 2010).

Independentemente da metodologia adotada, os guias disponíveis na literatura para uma boa prática de QSAR destacam que esse tipo de análise deve seguir ao menos cinco etapas:

- 1 – Seleção de um subgrupo de compostos químicos ou biológicos a ser analisado;
- 2 – Geração de descritores potentes que sejam capazes de refletir a essência estrutural dos compostos;
- 3 – Seleção dos descritores que serão incluídos no modelo de predição;
- 4 – Construção do modelo de calibração;
- 5 – Avaliação da estabilidade e validade do modelo gerado.

A depender do método QSAR, a estrutura de uma molécula (por exemplo, propriedades geométricas, eletrônicas e estéricas) deve apresentar informações importantes relacionadas com o comportamento físico, químico e biológico da molécula, e também, à capacidade de capturar essas informações em um ou mais descritores numéricos (TODESCHINI et al., 2020).

O sucesso no desenvolvimento da modelagem QSAR depende da seleção do conjunto de dados, dos descritores moleculares selecionados e da capacidade de previsão da atividade biológica. Quando os dados de atividade não forem suficientes para extrair padrões, os QSARs não terão um bom desempenho (LEELANANDA; LINDERT, 2016). Portanto, qualquer erro presente na estrutura resulta na incapacidade do cálculo de descritores ou em descritores errôneos, ou seja, os modelos desenvolvidos apresentam-se como incompletos,

imprecisos ou restritos apenas a uma fração dos dados que estão disponíveis e, assim, os modelos serão considerados imprecisos (TROPISHA, 2010).

Atualmente, os modelos de previsão que utilizam as técnicas QSAR podem contar apenas com as propriedades determinadas computacionalmente (descritores), sendo este um facilitador para o cálculo QSAR, minimizando o tempo de cálculo, permitindo a realização de previsões para uma maior quantidade de compostos e reduzindo os custos (TODESCHINI et al., 2020).

Os descritores são parâmetros necessários para correlacionar objetos (no presente caso, as moléculas) com as variáveis dependentes (as atividades biológicas) (FREITAS; BROWN; MARTINS, 2005). Os descritores mais clássicos representam a lipofilicidade das moléculas (por meio do coeficiente de partição octanol-água), além de propriedades e índices que informam sobre o número e tipo de átomos em uma molécula, bem como a conectividade entre eles. Por sua vez, descritores baseados em campos moleculares, que descrevem efeitos estéricos, eletrostáticos e forças inter/intramoleculares, têm se tornado bastante populares (ROY; KAR; DAS, 2015b). De forma geral, os descritores moleculares numéricos captam e convertem, por via teórica, as informações codificadas na estrutura molecular e podem ser utilizados para estabelecer relações quantitativas entre as estruturas químicas e suas respectivas propriedades, atividades biológicas e/ou outras propriedades experimentais. Em outras palavras, são representações matemáticas de uma molécula obtida por um algoritmo aplicado a uma representação molecular definida ou um procedimento experimental específico (TODESCHINI et al., 2020).

Ainda, de acordo com Todeschini et al. (2020):

Descritores moleculares são índices numéricos que codificam algumas informações relacionadas à estrutura molecular. Podem ser propriedades físico-químicas experimentais de moléculas e índices teóricos calculados por fórmulas matemáticas ou algoritmos computacionais. Os descritores moleculares estão intimamente ligados à estrutura molecular e desempenham um papel fundamental na pesquisa científica, constituindo o núcleo teórico de uma complexa rede de conhecimento (TODESCHINI et al., 2020).

Quanto à geração dos descritores moleculares teóricos, existem diversas abordagens que podem ser empregadas, sendo que cada uma dessas metodologias apresenta suas vantagens e desvantagens particulares. O QSAR-3D (leva em consideração as estruturas químicas tridimensionais para gerar os descritores), por exemplo, é uma das metodologias mais empregadas para geração de descritores; no entanto, apesar de gerar modelos confiáveis e interpretáveis, a sua manipulação é complexa, visto que se faz necessária uma exaustiva

busca conformacional e o alinhamento de estruturas tridimensionais de acordo com regras pré-estabelecidas (NUNES; FREITAS, 2013). Por outro lado, abordagens que se baseiam na utilização de descritores físico-químicos (abordagens 2D) para codificar mudanças estruturais são simples e geram modelos de predição confiáveis, mas, muitas vezes, não oferecem modelos facilmente racionalizáveis.

Portanto, uma abordagem que consiga balancear simplicidade, eficácia e interpretabilidade é considerada ideal. Nesse sentido, a metodologia MIA-QSAR (*Multivariate Image Analysis applied to QSAR*) (BARIGYE; FREITAS, 2016) parece ser uma interessante abordagem para modelagem molecular.

3.6 MIA-QSAR

MIA-QSAR (*Multivariate Image Analysis applied to QSAR*) é um método que se baseia no tratamento de imagens bidimensionais (2D) das estruturas químicas para modelagem molecular (BARIGYE; FREITAS, 2015). Entende-se que tais imagens contêm informações topoquímicas e topoestruturais (informações acerca da forma da molécula, efeitos estéricos, centros estereogênicos, etc.) relevantes e úteis para modelagem de propriedades físico-químico-biológicas. Em suma, as informações de caráter topoquímico e topoestrutural estão codificadas na forma de valores binários correspondentes aos pixels.

Como uma imagem digital é composta integralmente por **pixels** (valores binários que constituem a menor unidade de uma imagem digital), seus respectivos valores se tornam as variáveis independentes (**descritores**) que serão utilizadas para a construção do modelo de predição.

A metodologia MIA-QSAR tem como essência a ideia de que, para uma série de imagens de moléculas congêneres, as mudanças nos substituintes gerarão variações na orientação (posição) dos pixels, o que, por sua vez, explica as variações na variável resposta (propriedade biológica) (BARIGYE; FREITAS, 2015). Em outras palavras, as diferenças observadas na propriedade entre uma série de compostos semelhantes, estão relacionadas aos diferentes grupos químicos que esses compostos possuem (DARÉ; RAMALHO; FREITAS, 2018).

Essa técnica foi introduzida na década de 1980 em estudos de análise espectral, sensoriamento remoto, análises clínicas, entre outras (GELADI et al., 1992). Tal método desempenhava um papel muito importante nas diversas formas de análises de dados por sua interpretação ocorrer de maneira visual (FREITAS; BROWN; MARTINS, 2005).

No método MIA-QSAR, as imagens são representações das estruturas químicas dos compostos a serem estudados, que podem ser desenhadas usando qualquer *software* apropriado, e os descritores (pixels) são usados para correlacionar as estruturas químicas com suas respectivas atividades biológicas (propriedade biológica de interesse) (FREITAS; BROWN; MARTINS, 2005). Em outras palavras, quando uma série congênere de compostos é estudada, a porção variável das estruturas (os substituintes) pode ser tratada multivariadamente, com a finalidade de estabelecer uma correlação entre a estrutura química e a bioatividade, ou seja, descritores MIA podem ser usados para fazer previsões úteis acerca das porções das moléculas que estão relacionadas com a variável de interesse (FREITAS et al., 2008).

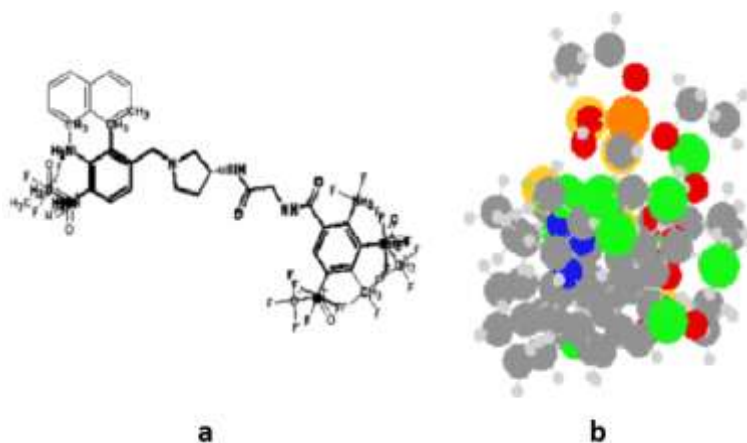
Na primeira versão do método MIA-QSAR, os átomos e ligações eram representados unicamente por contornos de cores preta e branca, ou seja, as moléculas consistiam em simples *wireframes* (FREITAS; BROWN; MARTINS, 2005). As variáveis decorrentes eram compostas pelos valores 0 (preto) e 765 (branco), no qual, 765 representava os espaços em branco da imagem, enquanto 0 indicava os traços e símbolos dos elementos químicos que compunham as estruturas químicas (BARIGYE; FREITAS, 2016). Apesar dessa metodologia conhecida como MIA-QSAR tradicional ter alcançado resultados promissores em diversas análises que podem ser encontradas na literatura, possuía algumas limitações em função do alinhamento impreciso dos *wireframes*, são elas:

- 1 - A imprecisão no alinhamento gerava variações nas posições dos pixels, o que, por sua vez, influenciava no modelo de regressão, ou seja, a dificuldade na etapa de alinhamento se tornava uma fonte de erro sistemático para a análise; o alinhamento ótimo depende, sobretudo, da precisão ocular e uma mudança mínima em alguns pixels pode distorcer significativamente os resultados do modelo (BARIGYE; FREITAS, 2015);
- 2 - Não considerava a diferença de tamanho dos átomos, o que fazia com que esse método falhasse em diferenciar heteroátomos e/ou tipos de átomos (NUNES; FREITAS, 2013);
- 3 - Codificava apenas informações topológicas e não havia incorporação de informações adicionais além dos *wireframes* para distinguir os diferentes átomos e/ou subestruturas nas imagens em preto e branco (BARIGYE; FREITAS, 2016).

Dessa forma, foi desenvolvida uma nova versão denominada *aug-MIA-QSAR* (*augmented MIA-QSPR*) com o intuito de melhorar a sua aplicabilidade e interpretabilidade, a qual se baseia nas cores oriundas do sistema RGB (*Red, Green and Blue*) e considera os diferentes raios atômicos na molécula (FIGURA 8). Essa nova abordagem apresentou-se como um marco importante na modelagem baseada em imagens multivariadas, pois foi

possível, pela primeira vez, incluir esquemas de ponderação para distinguir os diferentes átomos usando propriedades atômicas (NUNES; FREITAS, 2013).

Figura 8 — (a) Sobreposição de imagens utilizadas no modelo MIA-QSAR original, baseado em wireframes; (b) sobreposição de imagens utilizadas no modelo aug-MIA-QSAR.



Fonte: (a) (NUNES; FREITAS, 2013); (b) (DARÉ; SILVA; FREITAS, 2017).

Em aug-MIA-QSAR, os átomos são representados por círculos de tamanhos proporcionais aos respectivos raios de van der Waals e suas cores (valores de pixels) podem ser proporcionais a propriedades atômicas, tais como a eletronegatividade (FREITAS; BARIGYE; FREITAS, 2015). A criação dessa extensão aug-MIA-QSAR demonstrou uma melhora no poder preditivo e na interpretabilidade dos modelos MIA-QSAR, uma vez que informações importantes sobre as características estruturais mais relevantes para a propriedade de interesse eram obtidas, proporcionando maior facilidade no desenvolvimento de novos compostos químicos com propriedades desejadas (BARIGYE; FREITAS, 2015).

O esquema de cores empregue é descrito pelo sistema RGB, no qual, o espectro de cores é considerado um produto da contribuição das três cores primárias: vermelho, verde e azul, sendo que cada canal individual pode variar de intensidade de 0 a 255 (FREITAS; BARIGYE; FREITAS, 2015). Como um único pixel tem a contribuição das três cores, então, ele poderá assumir valores que vão de 0 (ausência de cor - preto) a 765 (contribuição das 3 cores - branco). Cada cor no sistema RGB-8 *bits* é representada por 1 *byte* de memória (8 *bits*), o que permite 256 valores diferentes.

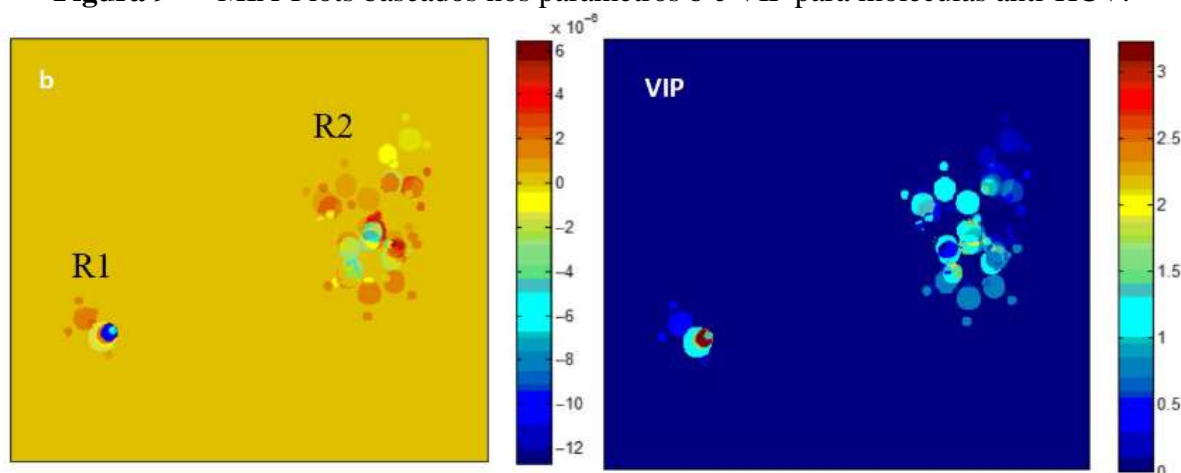
O uso de cores e raios atômicos para diferenciar átomos e grupos químicos fez com que o método aug-MIA-QSAR melhorasse sua habilidade de predição quando comparado com a metodologia convencional (NUNES; FREITAS, 2013).

Novos melhoramentos da técnica MIA-QSPR ocorreram em 2016, com o desenvolvimento de uma ferramenta gráfica: o MIA-Plot (BARIGYE et al., 2016). Essa ferramenta auxilia na interpretação química dos modelos gerados, ou seja, ela contribui para o entendimento de como os diferentes substituintes e suas respectivas posições afetam a propriedade biológica de interesse.

A necessidade do desenvolvimento de tal ferramenta surgiu devido ao fato de que o grande número de descritores gerados a partir dessa análise (geralmente milhares de pixels), fazia com que a interpretação química dos modelos MIA-QSAR fosse muito complicada. Assim, a **seleção de variáveis** foi utilizada para reduzir significativamente a quantidade de descritores, obtendo-se uma interpretação química dos modelos com base em uma menor quantidade de descritores (BARIGYE et al., 2016). Em termos práticos, essa nova ferramenta consiste na utilização de dois gráficos de cores, o primeiro é chamado de gráfico de coeficientes de regressão do PLS (**b**) e o segundo é o gráfico importância das variáveis em projeção para o grupo de compostos (VIP).

A importância dos descritores na determinação do modelo de projeção PLS é apresentada no gráfico de *scores* VIP, tanto para as variáveis preditoras quanto para a variável resposta, ou seja, ele permite a visualização da importância das variáveis para o modelo. Em contrapartida, um coeficiente de regressão indica como uma variável afeta a resposta, *i.e.*, um valor alto de **b** (positivo ou negativo) indica que tal variável tem uma maior relevância (correlação) para a propriedade modelada. Desse modo, recomenda-se uma análise conjunta dos gráficos VIP e **b**. A Figura 9 ilustra um exemplo de utilização dos MIA-Plots.

Figura 9 — MIA-Plots baseados nos parâmetros **b** e VIP para moléculas anti-HCV.



Fonte: (DARÉ; RAMALHO; FREITAS, 2018).

Para o desenvolvimento do presente trabalho, a técnica MIA-QSAR será utilizada para modelagem das propriedades biológicas das sulfonilureias propostas, enquanto a ferramenta MIA-Plot será aplicada para auxiliar na interpretação química e biológica dos resultados.

3.7 Validação de modelos QSAR

O procedimento de validação de um modelo QSAR consiste na avaliação do desempenho de ajuste e na avaliação da capacidade de predição do modelo. A capacidade preditiva dos modelos QSAR é considerada a característica mais importante para que um modelo seja aceitável. A avaliação da capacidade preditiva de um modelo é realizada utilizando-se um subconjunto de dados disponíveis como o conjunto de treinamento (para construção do modelo), e o conjunto de teste (consiste de 25 – 30 % do conjunto de dados original - é utilizado para a avaliação preditiva do modelo). A partição em conjuntos de treinamento / teste pode ser realizada de diferentes maneiras, dependendo do procedimento de validação (TODESCHINI et al., 2020).

Conjuntos de treinamento são usados no desenvolvimento de modelos e, quando muito pequenos, podem apresentar alguns problemas, tais como a chance de correlação ao acaso e o sobreajuste, não permitindo a construção de modelos verdadeiramente preditivos (TROPISHA, 2010).

O conjunto teste pode ser escolhido utilizando algum método estatístico (Kennard-Stone) ou pode ser escolhido aleatoriamente, de maneira que este seja representativo do conjunto original. O primeiro escolhe as amostras teste com base nas diferenças dos descritores, maximizando a distância Euclideana com base no número de componentes principais dos objetos, enquanto o segundo se baseia na distribuição das variáveis dependentes (atividades biológicas) das amostras.

Modelos provenientes de pequenos conjuntos de dados não podem ser testados em toda a sua extensão, podendo apresentar falhas e comportamentos anômalos em alguns dos testes de validação como, por exemplo, falta de robustez do modelo e na validação cruzada, sendo este um dificultador dos métodos de validação (KIRALJ; FERREIRA, 2009). Neste caso, o grupo de teste deve ser definido, levando-se em consideração a similaridade a alguns produtos químicos do conjunto de treinamento. Portanto, esse deve ser um grupo representativo de todos os compostos presentes no modelo original, de modo que não apresente grande semelhança e resulte em dados preditivos errôneos (TODESCHINI et al., 2020).

Existem diversos descritores disponíveis para descrever uma molécula e a dificuldade na identificação de quais características moleculares são as responsáveis por uma propriedade específica pode afetar drasticamente a previsibilidade de um modelo. Com o propósito de facilitar sua compreensão, descritores moleculares que sejam redundantes e não relevantes podem ser excluídos. De acordo com isso, uma das técnicas de redução de variável mais populares é baseada na variância do descritor. Descritores que apresentam variância nula são constantes para todas as moléculas do conjunto de dados e, portanto, não codificam informações relevantes para discriminar entre diversas estruturas moleculares. No entanto, estas podem ser deletadas do modelo (TODESCHINI et al., 2020).

A técnica de regressão por mínimos quadrados parciais (do inglês *Partial Least Squares* – PLS) é a técnica mais comumente usada em análises de QSAR nos procedimentos de calibração dos modelos. Essa técnica é utilizada para prever dados mais realistas de atividade biológica, podendo ser chamada também de método de estrutura latente ou de projeção. Segundo o método de regressão PLS, uma matriz de dados (X) é correlacionada com o vetor da função resposta (y) em uma combinação linear e por meio das componentes principais. As componentes principais são otimizadas para melhor descrever a relação entre os blocos X e y simultaneamente, ou seja, uma grande quantidade de descritores podem ser transformados em um número pequeno de novos termos ortogonais, chamados de variáveis latentes (LV). O número de variáveis latentes é usado para definir a variável dependente. Essa técnica tem como objetivo manter uma relação entre matrizes (características e propriedades) e, também, reduzir a dimensionalidade dos dados originais para o número de LV do modelo que apresentar o menor resíduo da variância de y (DAMALE et al., 2014).

O número de variáveis latentes ótimo é escolhido analisando-se o decaimento do RMSE (erro quadrático médio, do inglês *Root Mean Square Error*) na validação cruzada do tipo *Leave-One-Out*. O uso excessivo de variáveis latentes remete à tentativa de explicação de dados aleatórios, como ruídos. Portanto, somente um número razoavelmente pequeno de LV's é geralmente necessário. Para um número de variáveis latentes superior ao ideal, $RMSE_C$ e $RMSE_{CV}$ não são melhorados significativamente, e o uso de uma LV maior pode levar à realização de uma calibração excessiva do modelo, ou seja, um superajuste do modelo (*overfitting*).

As técnicas de validação são empregadas para garantir que a previsão de um modelo é confiável. Dentre as técnicas de validações disponíveis, uma das mais populares é a validação interna do tipo validação cruzada *leave-one-out* (LOO_{CV} , avaliada por meio do coeficiente de determinação q^2) e a validação externa (r^2_{pred} e RMSE correspondente). Pela técnica LOO_{CV} ,

cada molécula é sequencialmente deixada de fora uma vez, um modelo parcial é calibrado usando-se o conjunto de moléculas restante, e a previsão para a molécula deixada de fora é calculada usando os parâmetros de regressão obtidos no modelo parcial. No entanto, observa-se que para um número muito grande de amostras, o desempenho de previsão tende ao ajuste do modelo, proporcionando resultados excessivamente otimistas (TODESCHINI et al., 2020). Além disso, a proximidade entre os valores experimentais de atividade biológica e os valores calculados para o grupo pode ser avaliada estatisticamente utilizando o parâmetro r_m^2 (ROY et al., 2013):

$$r_m^2 = r^2 \times [1 - (r^2 - r_0^2)^{1/2}] \quad (1)$$

em que r^2 e r_0^2 são os coeficientes de determinação entre as atividades biológicas observadas e previstas com e sem intercepto, respectivamente.

O risco de correlação ao acaso (*chance correlation*) pode ser analisado por meio do parâmetro ${}^c r_p^2$ (MITRA; SAHA; ROY, 2010):

$$[{}^c r_p^2 = r \times (r^2 - r_{y\text{-random}}^2)^{1/2}] \quad (2)$$

em que $r_{y\text{-random}}^2$ corresponde ao valor do coeficiente de determinação médio obtido depois de randomizar o bloco y várias vezes. O $r_{y\text{-rand}}^2$ obtido pelo teste de *y-randomization* (coeficiente de determinação obtido após a randomização do bloco das variáveis dependentes) tem como objetivo detectar e quantificar as correlações aleatórias entre a variável dependente (y) e os descritores (x). Esse teste consiste na randomização do bloco y em várias execuções, mantendo-se a matriz X original fixa. Os modelos obtidos nessas condições devem ser de baixa qualidade e sem significado real; a presença de bons resultados indica que a correlação foi obtida ao acaso ou que o modelo está superajustado (*overfitting*) (KIRALJ; FERREIRA, 2009).

Outra técnica muito utilizada é a validação de *bootstrapping*. Essa técnica consiste na construção do conjunto de treinamento diversas vezes, incluindo o mesmo número inicial (n) de amostras, mas selecionando um novo conjunto de teste. Esse procedimento permite construir modelos a partir de conjuntos de treinamento sempre com o mesmo tamanho e é útil para avaliar a estabilidade do modelo. Além disso, essa técnica de validação tem como objetivo tornar os conjuntos de teste e treinamento tão representativos quanto possível (KIRALJ; FERREIRA, 2009).

Para a realização do cálculo de *bootstrapping*, seleciona-se um novo conjunto teste e mantém-se o número inicial de amostras, os conjuntos testes são selecionados de 15 a 25 vezes (15-25 ciclos) para a construção de novos modelos e, então, os resultados de r^2_{cal} , $r^2_{\text{LOOCV}} (q^2)$ e r^2_{m} são calculados a partir da média dos resultados obtidos para cada modelo construído. Modelos provenientes de pequenos conjuntos de dados dificilmente atendem todos os quesitos para um modelo considerado preditivo, pois esses podem apresentar falhas e comportamentos anômalos em alguns dos testes de validação, por exemplo, falta de robustez do modelo e na validação cruzada, sendo este um dificultador dos métodos de validação.

Ao contrário do *bootstrapping*, o teste de validação externa requer apenas uma divisão do conjunto de dados original em conjunto de treinamento estruturalmente semelhante e conjunto de teste, tendo como objetivo testar o poder preditivo do modelo (KIRALJ; FERREIRA, 2009).

A validação externa pode ser realizada com base em várias técnicas, entre elas, Coeficiente de Correlação de Concordância (do inglês *Concordance Correlation Coefficient* – CCC), Erro Médio Absoluto (do inglês *Mean Absolute Error* – MAE), k , k' e coeficientes de correlação quadrática preditiva (do inglês *Predictive Squared Correlation Coefficients* – Q^2_{F1} , Q^2_{F2} e Q^2_{F3}).

3.7.1 Coeficiente de Correlação de Concordância

A validação externa baseada no Coeficiente de Correlação de Concordância – CCC ou $\hat{\rho}_c$ (Fórmula 3) é calculada a partir da diferença entre os dados experimentais e os previstos (pelo modelo) do conjunto de teste (LIN, 1989). Essa técnica é mais intuitiva e conceitualmente mais simples, sendo capaz de avaliar o desempenho do modelo em termos de mensuração relativa dos erros. É uma técnica adequada para medir a concordância entre dados experimentais e previstos (CHIRICO; GRAMATICA, 2011).

$$CCC = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2} \quad (3)$$

em que (x_i) corresponde aos valores experimentais e (\bar{x}) corresponde à média dos valores do grupo teste, (y_i) corresponde aos valores preditos e (\bar{y}) corresponde à média dos valores preditos para o grupo teste, e (n) corresponde ao número de amostras do conjunto teste.

Esse coeficiente mede a precisão (distância entre os valores experimentais e os valores preditos) e a exatidão (distância entre as linhas de regressão e de inclinação passando pela

origem, ou seja, a linha de concordância). Como consequência, qualquer divergência da linha de regressão com a linha de concordância apresenta como resultado um valor de CCC menor que 1. Então, os dados obtidos correspondem perfeitamente a qualquer relação linear. Entretanto, neste último caso, o modelo pode ser preciso, mas não exato (CHIRICO; GRAMATICA, 2011).

O método de validação CCC apresenta algumas vantagens quando comparados a outros métodos. A primeira vantagem é a sua simplicidade e a sua maior independência da disposição dos eixos, enquanto a segunda vantagem é que nenhuma informação do conjunto de treinamento está envolvida e, então, ela pode ser considerada uma verdadeira medida de validação externa, independente do espaço químico amostrado (CHIRICO; GRAMATICA, 2011).

O valor de CCC não é tão restritivo, podendo ser escolhido arbitrariamente. Neste caso, o valor de corte proposto será de 0,85. Este valor é escolhido com base em um nível aceitável de espalhamento de dados de, no máximo, 0,15 (LIN, 1989). Os limites de corte são geralmente usados para cada medida de validação. Eles são aplicados em todos os conjuntos de dados simulados e, em seguida, para cada critério de validação. Em seguida, é realizada uma contagem dos modelos aceitos e rejeitados, com a finalidade de comparar quais dos modelos citados são considerados eficazes (CHIRICO; GRAMATICA, 2011).

3.7.2 Erro Médio Absoluto

A validação externa baseada no Erro Médio Absoluto – MAE (Fórmula 4), é considerada um determinante mais simples e direto dos erros de predição (ROY et al., 2016).

$$MAE = \frac{\sum_{i=1}^{n_{EXT}} |y_i - \hat{y}_i|}{n_{EXT}} \quad (4)$$

em que (y_i) são os valores experimentais do conjunto teste, (\hat{y}_i) são os valores preditos do conjunto teste e (n_{EXT}) é número de amostras do conjunto teste.

A validação MAE julga a dispersão dos erros e faz a comparação dos valores das métricas baseadas em erro após a remoção de uma fração infinita pequena (5%) de produtos químicos que possuem altos valores residuais para esse conjunto de teste.

De acordo com os valores de respostas obtidos, temos então:

1. Boas previsões - um erro de 10% da faixa do conjunto de treinamento deve ser aceitável, enquanto um valor de erro superior a 20% da faixa do conjunto de treinamento deve ser um erro muito alto.

2. Predições ruins - um valor de MAE superior a 15% da faixa do conjunto de treinamento deve ser alto, enquanto um valor de MAE superior a 25% da faixa do conjunto de treinamento é considerado muito alto.

As previsões que não se enquadrarem em nenhuma das condições acima podem ser consideradas de qualidade moderada. Os critérios acima devem ser aplicados para julgar a qualidade das previsões do conjunto de teste quando o número de dados for pelo menos 10 (confiabilidade estatística) e não houver erro sistemático nas previsões do modelo (aplicabilidade estatística) (ROY et al., 2016).

3.7.3 k e k'

Um modelo QSAR pode apresentar alta capacidade preditiva se estiver próximo do ideal. Isso implica que o valor do coeficiente de correlação (r^2) deve ser próximo de 1, levando-se em consideração os dados experimentais e preditos para um conjunto de dados. Os valores de k e k' devem ser próximos de 1 para condizer com os dados acima citados (GOLBRAIKH; TROPSHA, 2002).

A primeira formulação aqui descrita é a do k (Fórmula 5).

$$k = \frac{\sum(Y_{\text{obs}} \times Y_{\text{pred}})}{\sum(Y_{\text{pred}})^2} \quad (5)$$

em que (Y_{obs}) são os valores de atividades experimentais para os conjuntos de teste e/ou treinamento e (Y_{pred}) são os valores de atividades preditas para os conjuntos teste e/ou treinamento.

A segunda formulação aqui descrita é a do k' (Fórmula 6).

$$k' = \frac{\sum(Y_{\text{obs}} \times Y_{\text{pred}})}{\sum(Y_{\text{obs}})^2} \quad (6)$$

em que (Y_{obs}) são os valores de atividades experimentais para os conjuntos de teste e/ou treinamento e (Y_{pred}) são os valores de atividades preditas para os conjuntos teste e/ou treinamento.

Ambas as fórmulas se diferem no denominador: a primeira refere-se ao cálculo do somatório das atividades biológicas preditas elevadas ao quadrado, enquanto a segunda refere-se ao somatório das atividades biológicas experimentais elevadas ao quadrado.

De acordo com a literatura (ROY et al., 2013), para um modelo ser considerável aceitável, os valores de k e k' devem satisfazer o seguinte quesito: $0,85 \leq k \leq 1,15$ ou $0,85 \leq k' \leq 1,15$.

3.7.4 Q_{F1}^2 , Q_{F2}^2 e Q_{F3}^2

A avaliação do desempenho preditivo em conjuntos externos pode ser prevista pela validação designada *Predictive Squared Correlation Coefficients* - Q_{F1}^2 , Q_{F2}^2 e Q_{F3}^2 (CHIRICO; GRAMATICA, 2011).

A primeira formulação aqui designada é a de Q_{F1}^2 , descrita a seguir (Fórmula 7):

$$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y}_{TR})^2} \quad (7)$$

em que (y_i) são os valores experimentais do conjunto teste, (\hat{y}_i) são os valores preditos do conjunto teste e (\bar{y}_{TR}) é a média dos valores experimentais do conjunto treinamento.

Essa fórmula usa a média dos valores do conjunto de treinamento, para acompanhar a “distância” entre os dois conjuntos (teste e treinamento).

A segunda formulação é um método alternativo de cálculo (Fórmula 8).

$$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y}_{EXT})^2} \quad (8)$$

em que (y_i) são os valores experimentais do conjunto teste, (\hat{y}_i) são os valores preditos do conjunto teste e (\bar{y}_{EXT}) é a média dos valores experimentais do conjunto teste.

Essa fórmula (Fórmula 8) se difere da anterior (Fórmula 7) pelo fato de o valor médio no denominador ser calculado usando os dados experimentais do conjunto de teste, e não, os dados do conjunto de treinamento. Esse método independe dos valores de “distância” média de treinamento.

Uma nova formulação foi proposta (Fórmula 9), o critério de validação Q_{F3}^2 , que é uma alternativa para superar as desvantagens de se trabalhar com uma distribuição de dados tendenciosos (apenas com o conjunto teste ou apenas com dados médios do conjunto treinamento).

$$Q_{F3}^2 = 1 - \frac{[\sum_{i=1}^{n_{EXT}} (\hat{y}_i - y_i)^2] / n_{EXT}}{[\sum_{i=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2] / n_{TR}} \quad (9)$$

em que (n_{EXT}) é o número de amostras do conjunto teste e (n_{TR}) é o número de amostras do conjunto treinamento.

Como pode ser visto, esta formulação difere de Q_{F1}^2 e Q_{F2}^2 , pois o denominador é calculado no conjunto de treinamento e ambos numerador e denominador são divididos pelo número dos elementos correspondentes.

As previsões externas para Q_{F1}^2 , Q_{F2}^2 e Q_{F3}^2 são consideradas boas quando o limite usado apresenta um valor maior que 0,6 (CHIRICO; GRAMATICA, 2011).

3.8 Docking molecular

A proposição de novos compostos candidatos a fármacos e/ou agroquímicos é baseada em abordagens envolvendo ligantes e receptores. Esta última consiste na determinação da afinidade existente entre o substrato e a proteína e, portanto, é necessária uma etapa designada *docking* molecular para compreender as interações presentes entre o ligante (sulfonilureia) e o receptor (sítio ativo da enzima) (PAGADALA; SYED; TUSZYNSKI, 2017).

Essa ferramenta computacional é amplamente utilizada, por se apresentar como uma ferramenta ágil e econômica, sendo assim, empregada na predição *in silico* dos modos de interação e afinidades envolvidos nos eventos de reconhecimento molecular (DU et al., 2016; THOMSEN; CHRISTENSEN, 2006).

Os programas utilizados para a realização de estudos de *docking*, tais como o *Molegro Virtual Docker* (MVD), são capazes de calcular a interação ligante-receptor através da otimização da estrutura tridimensional de uma molécula pequena (ligante) no sítio de ligação do receptor e, as moléculas que apresentarem menor energia recebem uma pontuação (*score*), a qual indica a probabilidade de ligação com o receptor (CAVASOTTO; AUCAR; ADLER, 2018).

A pontuação é usada para avaliar as interações entre a proteína e o ligante, além de classificar os complexos alvo-ligante previstos por algoritmos de encaixe. As funções de pontuação são métodos matemáticos aproximados e rápidos utilizados para avaliar a afinidade de ligação entre a proteína e o ligante após o acoplamento (DU et al., 2016). A função de pontuação usada por algoritmos de encaixe é uma parte crucial do algoritmo, podendo ser baseadas na combinação de funções empíricas com funções geométricas e energéticas, tais

como comprimentos de ligação, ângulos de ligação, forças de Van der Waals, Lenard-Jones, interações eletrostáticas, entre outras (LEELANANDA; LINDERT, 2016; SALES, 2018).

A Equação 10 define os valores da função de pontuação.

$$E_{score} = E_{inter} + E_{intra} \quad (10)$$

em que E_{inter} corresponde a interação que ocorre entre o ligante e a proteína, e E_{intra} representa a energia interna do ligante.

A metodologia de *docking* molecular visa prever os modos de ligação entre o ligante e seus alvos biomoleculares de interesse com base na hipótese de que uma pontuação de energia mais baixa representa uma melhor postura de ligação proteína-ligante, ajudando a elucidar os principais elementos que determinam as interações proteína-ligante (GAJO et al., 2017).

A modelagem de uma proteína que não está presente em bases de dados pode ser realizada por homologia, a qual se baseia na semelhança entre as estruturas primárias dessa proteína e de proteínas homólogas de estruturas tridimensionais conhecidas (proteínas-molde), implicando na similaridade estrutural entre elas. A modelagem por homologia é, atualmente, o método computacional mais preciso para gerar modelos estruturais confiáveis e é usado rotineiramente em muitas aplicações biológicas. Tal metodologia tem como objetivo construir modelos tridimensionais de estruturas de proteínas usando estruturas determinadas experimentalmente de modelos relacionados a eles (WATERHOUSE et al., 2018).

Para uma determinada proteína alvo, uma biblioteca de estruturas de proteínas experimentais é pesquisada para identificar modelos adequados. Um modelo tridimensional para a proteína alvo será gerada com base em um alinhamento de sequência entre a proteína alvo e a estrutura modelo. As ferramentas de avaliação da qualidade do modelo são usadas para estimar a confiabilidade dos modelos resultantes (WATERHOUSE et al., 2018).

Juntamente com as metodologias QSAR, os estudos de *docking* auxiliam na validação e na interpretação da orientação e as conformações no sítio de ligação, sendo essenciais para a compreensão do mecanismo de interação entre o ligante (compostos bioativos) com o seu alvo biológico (MONTANARI, 2011).

4 METODOLOGIA

Um grupo de 23 compostos análogos pertencentes à classe das sulfonilureias, bem como os respectivos dados biológicos descritos em termos de pEC₅₀ (EC₅₀ em mol L⁻¹), porcentagem de atividade antifúngica e porcentagem de atividade herbicida, foram obtidos na literatura (WEI et al., 2016a).

A Tabela 2 apresenta os dados de atividade antifúngica obtidos da literatura (WEI et al., 2016a) referentes aos 23 compostos análogos pertencentes à classe das sulfonilureias, bem como os respectivos dados biológicos descritos em termos de pEC₅₀ (*C. cornigerum*) e porcentagem de atividade antifúngica.

Tabela 2 — Análogos de sulfonilureias e suas respectivas atividades antifúngicas.

Compostos	pEC ₅₀ <i>C. cornigerum</i>	% A. <i>solani</i>	% S. <i>sclerotiorum</i>	% C. <i>arachidicola</i>	% P. <i>piricola</i>	% C. <i>cornigerum</i>	% C. <i>lagenarium</i>
9 ^a	1,655	73,3	71,4	57,1	83,0	68,9	45,5
9 ^b	1,644	46,7	57,1	57,1	63,8	72,1	36,4
9 ^c	1,596	40,0	10,7	52,4	72,3	75,4	36,4
9 ^d	1,612	40,0	10,7	61,9	68,1	72,1	63,6
9 ^e	1,688	40,0	67,9	66,7	74,5	73,8	51,5
9 ^f	1,644	46,7	53,6	66,7	61,7	72,1	36,4
9 ^g	1,758	83,3	71,4	57,1	85,1	80,3	51,5
9 ^h	1,775	40,0	66,1	66,7	76,6	82,0	33,3
9 ⁱ	1,651	46,7	66,1	66,7	70,2	78,7	51,5
9 ^j	1,865	46,7	69,6	61,9	74,5	78,7	57,6
9 ^k	1,877	66,7	64,3	47,6	66,0	72,1	57,6
9 ^l	1,743	40,0	10,7	61,9	76,6	72,1	36,4
9 ^m	1,818	33,3	60,7	66,7	87,2	73,8	66,7
9 ⁿ	1,830	26,7	42,9	66,7	91,5	73,8	45,5
9 ^o	1,751	26,7	10,7	71,4	70,2	77,0	51,5
9 ^p	1,903	46,7	67,9	71,4	66,0	82,0	48,5
9 ^q	1,860	63,3	60,7	81,0	76,6	83,6	78,5
9 ^r	1,063	40,0	25,0	19,0	2,1	16,4	27,3
9 ^s	1,154	40,0	71,4	42,9	14,9	24,6	27,3
9 ^t	1,763	31,3	88,0	75,0	51,2	75,0	88,5
9 ^u	1,592	31,3	46,0	45,0	51,2	66,1	69,2
9 ^v	1,428	6,3	33,3	33,3	16,1	66,7	23,5
9 ^w	1,756	25,0	62,5	66,7	41,9	84,8	58,8

Fonte: (WEI et al., 2016a).

A Tabela 3 apresenta os dados de atividade herbicida obtidos da literatura (WEI et al., 2016a) referentes aos 23 compostos análogos pertencentes à classe das sulfonilureias, bem como os respectivos dados biológicos descritos em termos de porcentagem de atividade herbicida expressos em termos de % de crescimento da planta, para aplicações pré e pós-emergente contra *Brassica napus* e *Amaranthus retroflexus*, e os valores médios dos efeitos herbicidas.

Tabela 3 — Análogos de sulfonilureias e suas respectivas atividades herbicidas.

Compostos	% <i>Brassica napus</i>			% <i>Amaranthus retroflexus</i>		
	pré-emergente	pós-emergente	Média	pré-emergente	pós-emergente	Média
9a	93,1	69,0	81,05	67,6	10,0	38,80
9b	88,0	67,1	77,55	95,4	10,0	52,70
9c	51,3	76,2	63,75	0,0	0,0	0,00
9d	64,8	25,8	45,30	70,7	0,0	35,35
9e	10,0	36,0	23,00	80,0	0,0	40,00
9f	90,9	92,5	91,70	67,1	98,4	82,75
9g	35,0	26,5	30,75	19,1	93,5	56,30
9h	45,7	20,1	32,90	0,0	0,0	0,00
9i	96,5	45,6	71,05	60,7	37,4	49,05
9j	51,0	44,4	47,70	0,0	37,9	18,95
9k	48,6	34,5	41,55	65,9	43,7	54,80
9l	43,8	39,8	41,80	71,7	79,1	75,40
9m	10,0	23,4	16,70	54,9	23,3	39,10
9n	0,0	21,1	10,55	5,0	68,2	36,60
9o	30,0	25,8	27,90	0,0	23,9	11,95
9p	20,0	25,0	22,50	0,0	61,9	30,95
9q	15,0	87,7	51,35	10,0	33,2	21,60
9r	15,0	89,9	52,45	15,0	85,6	50,30
9s	10,0	66,1	38,05	0,0	32,7	16,35
9t	98,0	97,0	97,50	99,0	85,6	92,30
9u	89,2	98,0	93,60	99,0	66,9	82,95
9v	99,1	85,9	92,50	98,8	80,6	89,70
9w	10,0	5,0	7,50	0,0	5,0	2,50

Fonte: (WEI et al., 2016a).

As porcentagens de atividade antifúngica referem-se a uma dosagem de 50 mg L⁻¹ e as porcentagens de atividade herbicida referem-se a uma dosagem de 375 g hm⁻², de acordo com a literatura (WEI et al., 2016a).

Os dados de atividades antifúngicas em EC_{50} (concentração necessária para induzir metade do efeito máximo, em mol L^{-1}) foram convertidos em pEC_{50} ($-\log EC_{50}$) e usados na modelagem MIA-QSAR, enquanto os dados descritos em valores de porcentagem foram convertidos em I (% de inibição na escala decimal), pI ($-\log I$) e unidade probit (FINNEY, 1952), para serem utilizados na modelagem MIA-QSAR.

Os dados de atividades herbicidas foram baseados na porcentagem média dos valores pré e pós-emergentes, que foram posteriormente convertidos em I (% de inibição na escala decimal), pI ($-\log I$) e unidade probit (FINNEY, 1952), para serem usados na modelagem MIA-QSAR.

O melhor desempenho de modelagem para as atividades antifúngica e herbicida foi encontrado com os dados de pEC_{50} contra *Ceratobasidium cornigerum* e I contra *Brassica napus* respectivamente, que foram utilizados para a obtenção dos índices de seletividade ($SI = pEC_{50} / I$). O conjunto de dados dos compostos e os respectivos valores biológicos usados nas análises MIA-QSAR são apresentados na Tabela 4.

Tabela 4 — Série de 23 sulfonilureias com atividade antifúngica contra *C. cornigerum* (pEC₅₀), atividade herbicida contra *B. napus* (*I*) e índice de seletividade (*SI*).

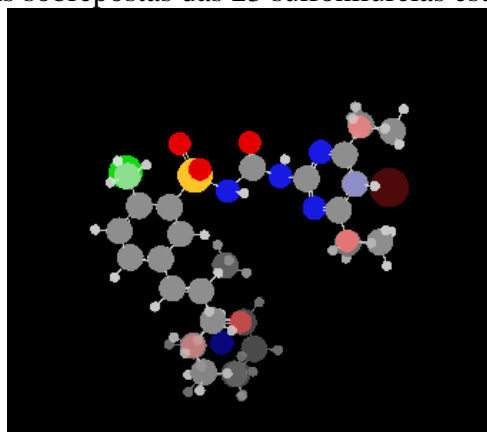
Composto	R ₁	R ₂	R ₃	R ₄	R ₅	X	pEC ₅₀	<i>I</i>	SI
1	CH ₃	H	H	OCH ₃	OCH ₃	C-H	1,655	0,811	2,04
2	CH ₃	H	CN	OCH ₃	OCH ₃	C-H	1,644	0,776	2,12
3	CH ₃	H	CN	CH ₃	CH ₃	C-H	1,596	0,638	2,50
4 ^a	CH ₃	H	COOCH ₃	OCH ₃	OCH ₃	C-H	1,612	0,453	3,56
5	CH ₃	H	COOCH ₃	CH ₃	CH ₃	C-H	1,688	0,230	7,34
6	CH ₃	H	CH ₃	OCH ₃	OCH ₃	C-H	1,644	0,917	1,79
7	CH ₃	H	H	CH ₃	CH ₃	C-H	1,758	0,308	5,72
8	CH ₃	H	H	H	CH ₃	C-H	1,775	0,329	5,40
9 ^a	CH ₃	H	CH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1,651	0,711	2,32
10 ^a	CH ₃	H	CH ₂ CH ₃	CH ₃	CH ₃	C-H	1,865	0,477	3,91
11	CH ₃	H	CH ₂ CH ₃	H	CH ₃	C-H	1,877	0,416	4,52
12	CH ₃	H	COOCH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1,743	0,418	4,17
13	CH ₃	H	COOCH ₂ CH ₃	CH ₃	CH ₃	C-H	1,818	0,167	10,89
14 ^a	CH ₃	H	COOCH ₂ CH ₃	H	CH ₃	C-H	1,830	0,106	17,35
15	CH ₃	H	CH ₂ CH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1,751	0,279	6,28
16	CH ₃	H	CH ₂ CH ₂ CH ₃	CH ₃	CH ₃	C-H	1,903	0,225	8,46
17	CH ₃	H	C ₆ H ₅	OCH ₃	OCH ₃	C-H	1,860	0,514	3,62
18	CH ₃	CH ₃	COOCH ₃	OCH ₃	OCH ₃	C-H	1,063	0,525	2,03
19	CH ₃	CH ₃	COOCH ₃	CH ₃	CH ₃	C-H	1,154	0,318	3,63
20	Cl	H	H	OCH ₃	OCH ₃	C-H	1,763	0,975	1,81
21	Cl	H	CH ₃	OCH ₃	OCH ₃	C-H	1,592	0,936	1,70
22	Cl	H	H	CH ₃	OCH ₃	N	1,428	0,925	1,54
23	Cl	H	H	CH ₃	OCH ₃	C-Br	1,756	0,075	23,41

^a Compostos de conjunto de teste (o mesmo usado em WEI et al., 2016a).

Fonte: Do autor, 2021.

As moléculas propostas neste estudo foram desenhadas no programa *GaussView* (DENNINGTON et al., 2008), alinhando-as de forma que o núcleo comum permanecesse na mesma posição, em termos de pixels (FIGURA 10), em todas elas, para garantir que, assim, a superposição dos compostos fosse perfeita nos pontos em comum. As imagens foram salvas como bitmaps com dimensões de 228 × 200 pixels.

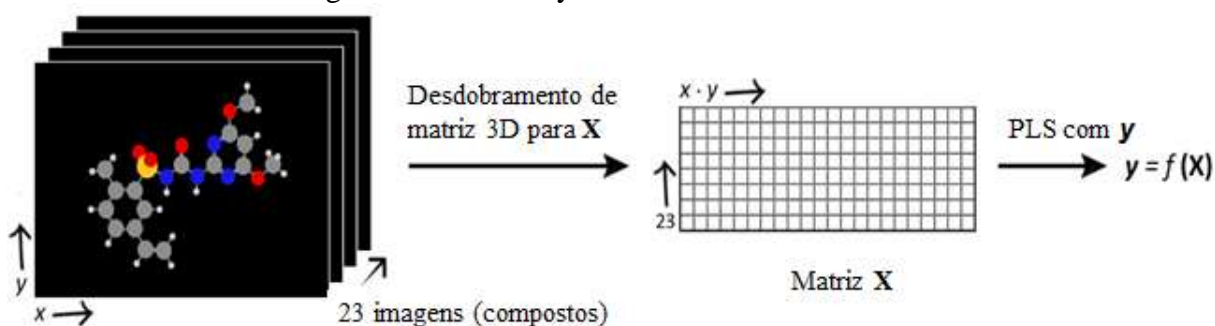
Figura 10 — Imagens sobrepostas das 23 sulfonilureias estudadas neste trabalho.



Fonte: Do autor, 2021.

Em suma, as imagens foram salvas com dimensões x e y definidas. Em seguida, elas foram sobrepostas para formar um arranjo tridimensional de dimensão $N \times x \times y$ (onde N corresponde ao número de moléculas). Esse arranjo 3-D foi desdobrado em uma matriz de dimensões $N \times (x \times y)$, na qual cada linha corresponde a uma molécula e cada coluna corresponde a um pixel de uma determinada posição da imagem; em outras palavras, a coluna 1, por exemplo, corresponde ao pixel de posição (1,1) de cada uma das imagens (FIGURA 11).

Figura 11 — Fluxo de trabalho MIA-QSAR: conversão de imagem em uma matriz de e descritores e regressão PLS com y .



Fonte: Do autor, 2021.

Posteriormente, as colunas com variância ‘0’ foram excluídas e, então, os valores das cores correspondentes a cada átomo foram substituídos por valores proporcionais ao raio de van der Waals do respectivo átomo (r_{vdw}), à eletronegatividade de Pauling do átomo (ϵ), e à razão r_{vdw}/ϵ (razão entre o raio de van der Waals e a eletronegatividade), originando as três matrizes de dados baseadas em imagens obtidas do *GaussView* para serem utilizadas na análise QSAR (TABELA 5).

Tabela 5 — Valores de pixel originais do programa GaussView e os respectivos valores proporcionais a propriedades atômicas pelos quais foram substituídos.

Átomos/ligação	Valores do pixel	r_{vdW}	ϵ	r_{vdW} / ϵ
Carbono	426	170	255	667
Hidrogênio	612	120	220	545
Nitrogênio	279	155	304	510
Oxigênio	229	152	344	442
Cloro	289	175	316	554
Bromo	231	185	296	625
Ligação	613	0,1	0,1	0,1

Fonte: Do autor, 2021.

Em seguida, deu-se início à análise QSAR, sendo gerados dois modelos para cada tipo de descritor MIA (r_{vdW} , ϵ e r_{vdW}/ϵ); o primeiro utilizou os dados de pEC₅₀ para a atividade antifúngica contra *C. cornigerum*, enquanto o segundo utilizou os dados de I para a atividade herbicida contra *B. napus*. Um teste de domínio de aplicabilidade usando *Sample Leverages vs. Studentized Residuals* foi realizado para avaliar a presença de possíveis *outliers* no conjunto de dados (ROY et al., 2016).

Posteriormente, cada matriz de descritores foi dividida em conjuntos de treinamento e teste, de acordo com a divisão relatada anteriormente na literatura, para fins de comparação (WEI et al., 2016a).

O procedimento de calibração foi realizado utilizando-se regressão por mínimos quadrados parciais (*Partial Least Squares* - PLS), sendo que o número de variáveis latentes foi escolhido analisando-se o decaimento do RMSE (erro quadrático médio, do inglês *Root Mean Square Error*) na validação cruzada do tipo *leave-one-out* por meio do programa Chemoface (NUNES et al., 2012).

A qualidade do modelo de predição foi avaliada por intermédio de parâmetros provenientes da validação interna feita por *leave-one-out* (q^2 e RMSE correspondente) e externa (r_m^2 e RMSE correspondente). Além disso, a proximidade entre os valores experimentais de atividade biológica e os calculados para o grupo teste foi avaliada estatisticamente utilizando o parâmetros r_m^2 , e o risco de correlação ao acaso (*chance correlation*) foi analisado por meio do parâmetro $^c r_p^2$. O domínio de aplicabilidade do modelo foi avaliado plotando-se o gráfico de *sample leverages* e *Student's residuals*. Outros métodos de validação externa também foram avaliados, quais sejam: *bootstrapping*, no qual 20 conjuntos de testes foram selecionados aleatoriamente, e as etapas de calibração e validação

foram realizadas para cada ciclo, sendo assim, coletados os valores estatísticos médios; CCC, que mede a precisão (distância entre os valores experimentais e os valores preditos) e a exatidão (distância entre as linhas de regressão e de inclinação passando pela origem, ou seja, a linha de concordância); MAE, que julga a dispersão dos erros e faz a comparação dos valores das métricas baseadas em erro após a remoção de uma fração pequena de produtos químicos que apresentem altos valores de resíduos para o conjunto de teste; k e k' , os quais estão relacionados com a inclinação da reta; Q^2_{F1} , Q^2_{F2} e Q^2_{F3} , que correspondem a coeficientes de correlação quadrática preditiva usados para avaliar o desempenho preditivo em conjuntos externos.

Posteriormente, para fins de interpretação, os mapas de contorno MIA foram obtidos plotando-se os coeficientes de regressão PLS (**b**) e os *scores* do gráfico VIP (*Variable Importance in Projection*). A partir dos gráficos MIA, foi possível analisar quais as características químicas que mais afetam as variáveis de resposta. O gráfico VIP indica quais são as porções das moléculas que afetam principalmente a variável de resposta, enquanto o gráfico **b** consiste em um mapa de contorno capaz de descrever os coeficientes de regressão PLS, ou seja, é capaz de identificar os substituintes que contribuem para aumentar ou diminuir as atividades.

Após a análise dos MIA-plots, novos candidatos a agroquímicos foram propostos, de acordo com as características químicas que influenciam as atividades antifúngica e herbicida.

Um estudo de *docking* molecular foi realizado com o intuito de explicar a afinidade das sulfonilureias à AHAS (EC 2.2.1.6). Para tal, foi utilizada a estrutura cristalizada de *Saccharomyces cerevisiae* aceto-hidroxiácido sintase (*ScAHAS*) complexada ao *sulfometuron-metil*, a uma resolução de 2,34 Å obtida a partir do PDB (*Protein Data Bank*). A estrutura de código PDB 1T9C com *sulfometuron-metil* foi usada devido à similaridade estrutural entre o inibidor cristalizado e os compostos utilizados neste estudo (FIGURA 12). Além disso, o *sulfometuron-metil* é um membro comercialmente importante da família dos herbicidas de sulfonilureia e que tem como alvo a AHAS. A estrutura cristalina de *B. napus* AHAS (*BnAHAS*) foi obtida por modelagem por homologia (FIGURA 13) (FISER, 2010).

Figura 12 – Estrutura cristalina de aceto-hidroxiácido sintase de levedura em complexada com um herbicida de sulfonilureia, sulfometuron metil.



Fonte: Protein Data Bank ((J. A. McCourt, S. S. Pang, L. W. Guddat, R. G. Duggleby, R.G., Crystal Structure Of Yeast Aceto-hydroxyacid Synthase In Complex With A Sulfonylurea Herbicide, Sulfometuron methyl, *Biochemistry*, 2005, **44**, 2330-2338.

Figura 13 – Estrutura cristalina de *B. napus* AHAS (*Bn*AHAS) obtida por modelagem por homologia.



Fonte: Do autor, 2021.

Para iniciar a modelagem, foi realizada uma busca sistemática no Swiss Bioinformatics Resource Portal (ExPasy) para obter a estrutura primária do *BnAHAS* usando o banco de dados UniProt (CONSORTIUM, 2020) como ferramenta de busca. O perfil obtido foi usado para a pesquisa de estruturas de proteínas conhecidas no PDB, usando a sequência alvo como consulta.

Posteriormente, a sequência primária da proteína alvo foi alinhada com a proteína modelo usando o programa de similaridade SIM (HUANG; MILLER, 1991). Os modelos foram classificados com base na pontuação de alinhamento e na qualidade estrutural de acordo com WHAT_CHECK (HOOFT et al., 1996). O *BnAHAS* foi construído usando o servidor SWISS-MODEL (WATERHOUSE et al., 2018).

As estruturas de todos os ligantes da Tabela 4 foram preparados e otimizados com parâmetros de campo de força AMBER (GAFF) e cargas AM1-BCC. Em seguida, os ligantes foram ancorados no local de ligação da AHAS usando o software *Molegro Virtual Docker* (THOMSEN; CHRISTENSEN, 2006), que é projetado para a previsão da conformação de ligação mais provável de um ligante em uma macromolécula (RAMALHO et al., 2010). As funções de pontuação (*score*) baseadas em grade pré-calculam os valores de energia potencial em uma grade cúbica uniformemente espaçada para acelerar os cálculos. A energia potencial foi avaliada usando interpolação tri-linear entre os pontos relevantes da grade. Foi usada uma grade de 0,30 Å de resolução e 12 Å de raio do inibidor como o local de ligação. MolDock SE foi usado como um algoritmo de busca e o número máximo de poses a serem geradas foi 100.

Por fim, o comportamento dos compostos pode ser interpretado por meio de estudos de *docking*, os quais ofereceram suporte aos resultados do QSAR.

5 CONSIDERAÇÕES GERAIS

As enzimas aceto-hidroxiácido sintase (AHAS) de fungos e plantas são alvos biológicos importantes de pesticidas à base de sulfonilureias, e a introdução de uma porção alquenila em uma série de sulfonilureias mostrou induzir seletivamente à atividade antifúngica sobre a atividade herbicida, podendo ser útil para a conservação de forragem. Portanto, as atividades antifúngica e herbicida dessa série de compostos foram modeladas usando MIA-QSAR, e o índice de seletividade decorrente da relação atividade antifúngica/herbicida foi aplicado para avaliar o potencial de aplicação dos novos agroquímicos propostos. Essas propostas foram concebidas com base nas características químicas que mais afetam positivamente as atividades antifúngicas e que desfavorecem as propriedades herbicidas. Esses efeitos provém, principalmente, de substituintes nas posições X (C-Br), R₃ (-COOEt, -Et e -Pr) e R₅ (Me). Como resultado da combinação dessas subestruturas, pelo menos 9 dos 46 novos candidatos a agroquímicos apresentaram índices de seletividade calculados muito promissores. Os estudos de docking em AHAS corroboraram com esses resultados; no entanto, os 9 compostos propostos mais promissores devem ser sintetizados e testados para legitimar os resultados computacionais.

REFERÊNCIAS

- AMORIM, D. S. et al. Sesame production and composition compared with conventional forages. **Chilean Journal of Agricultural Research**, v. 79, n. 4, p. 586-595, 2019.
- BARIGYE, S. J.; FREITAS, M. P. Ten Years of the MIA-QSAR Strategy. **International Journal of Quantitative Structure-Property Relationships**, v. 1, n. 1, p. 64-77, 2016.
- BARIGYE, S. J. et al. MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. **RSC Advances**, v. 6, n. 55, p. 49604-49612, 2016.
- BARIGYE, S. J.; FREITAS, M. P. 2D-Discrete Fourier Transform: generalization of the mia-qsar strategy in molecular modeling. **Chemometrics and Intelligent Laboratory Systems**, v. 143, p. 79-84, 2015.
- CAVASOTTO, C. N.; AUCAR, M. G.; ADLER, N. S. Computational chemistry in drug lead discovery and design. **International Journal of Quantum Chemistry**, v. 119, n. 2, p. 1-19, 2018.
- CHEN, W. et al. Synthesis and evaluation of novel N-(4'-arylpyrimidin-2'-yl) sulfonylurea derivatives as potential antifungal agents. **Chemical Research in Chinese Universities**, v. 31, n. 2, p. 218-223, 2015.
- CHEN, W. et al. Design, synthesis and SAR study of novel sulfonylurea derivatives containing arylpyrimidine moieties as potential anti-phytopathogenic fungal agents. **Chinese Chemical Letters**, v. 30, n. 12, p. 2160-2162, 2019.
- CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models: how to evaluate it? comparison of different validation criteria and proposal of using the concordance correlation coefficient. **Journal of Chemical Information and Modeling**, v. 51, n. 9, p. 2320-2335, 2011.
- CONSORTIUM, The Uniprot. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. 1, p. D480-D489, 2020.
- DAMALE, M. et al. Recent Advances in Multidimensional QSAR (4D-6D): a critical review. **Mini-Reviews in Medicinal Chemistry**, v. 14, n. 1, p. 35-55, 2014.
- DARÉ, J. K.; RAMALHO, T. C.; FREITAS, M. P. 3D perspective into MIA-QSAR: a case for anti-HCV agents. **Chemical Biology & Drug Design**, v. 93, n. 6, p. 1096-1104, 2018.
- DARÉ, J. K.; SILVA, C. F.; FREITAS, M. P. Revealing chemophoric sites in organophosphorus insecticides through the MIA-QSPR modeling of soil sorption data. **Ecotoxicology and Environmental Safety**, v. 144, p. 560-563, 2017.
- DENNINGTON, R. D.; KEITH, T. A.; MILLAM, M. J. GaussView 5.0, Wallingford, CT, 2008.

DU, X. et al. Insights into Protein–Ligand Interactions: mechanisms, models, and methods. **International Journal of Molecular Sciences**, v. 17, n. 2, p. 144-178, 2016.

DUGGLEBY, R. G.; MCCOURT, J. A.; GUDDAT, L. W.. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. **Plant Physiology and Biochemistry**, v. 46, n. 3, p. 309-324, 2008.

DUGGLEBY, R.; PANG, S. S. Acetohydroxyacid Synthase. **Journal of Biochemistry and Molecular Biology**, v.33, p. 1-36, 2000.

FINNEY, D. J. Probit Analysis. 2nd ed. **Journal of the American Pharmaceutical Association (Scientific Ed.)**, New York, v. 41, n. 11, p. 627, 1952.

FISER, A. Template-Based Protein Structure Modeling. **Methods in Molecular Biology**, p. 73-94, 2010.

FREITAS, M. P.; BROWN, S. D.; MARTINS, J. A. MIA-QSAR: a simple 2d image-based approach for quantitative structure-activity relationship analysis. **Journal of Molecular Structure**, v. 738, n. 1-3, p. 149-154, 2005.

FREITAS, M. P. et al. Multimode Methods Applied on MIA Descriptors in QSAR. **Current Computer Aided-Drug Design**, v. 4, n. 4, p. 273-282, 2008.

FREITAS, M. R.; BARIGYE, S. J.; FREITAS, M. P.. Coloured chemical image-based models for the prediction of soil sorption of herbicides. **RSC Advances**, v. 5, n. 10, p. 7547-7553, 2015.

GAJO, G. C. et al. Multi-objective Optimization of Benzamide Derivatives as Rho Kinase Inhibitors. **Molecular Informatics**, v. 37, n. 3, p. 1700080, 2017.

GARCIA, M. D. et al. Commercial AHAS-inhibiting herbicides are promising drug leads for the treatment of human fungal pathogenic infections. **Proceedings of the National Academy of Sciences**, v. 115, n. 41, p. E9649-E9658, 2018.

GELADI, P. et al. Image analysis in chemistry II. Multivariate image analysis. **Trac Trends in Analytical Chemistry**, v. 11, n. 3, p. 121-130, 1992.

GOLBRAIKH, A.; TROPSHA, A. Beware of q²! **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 269-276, 2002.

GOODE-ROMERO, G.; AGUAYO-ORTIZ, R.; DUEÑAS, L. D. Relaciones cuantitativas estructura-actividad/propiedad en dos dimensiones empleando el programa R. **Educación Química**, v. 30, n. 2, p. 27-40, 2019.

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

HAY, J. V. Chemistry of sulfonylurea herbicides. **Pesticide Science**, v. 29, n. 3, p. 247-261, 1990.

- HOOFT, R. W. W. et al. Errors in protein structures. **Nature**, v. 381, n. 6580, p. 272, 1996.
- HUANG, X.; MILLER, W. A time-efficient, linear-space local similarity algorithm. **Advances in Applied Mathematics**, v. 12, p. 337-357, 1991.
- IERICH, J. C. M. et al. A Computational Protein Structure Refinement of the Yeast Acetohydroxyacid Synthase. **Journal of the Brazilian Chemical Society**, v. 26, n. 8, p. 1702-1709, 2015.
- 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**, v. 20, n. 4, p. 770-787, 2009.
- KLEINSCHMIT, D.H.; SCHMIDT, R.J.; KUNG JUNIOR, L. The Effects of Various Antifungal Additives on the Fermentation and Aerobic Stability of Corn Silage. **Journal of Dairy Science**, v. 88, n. 6, p. 2130-2139, 2005.
- KUMAR, V. et al. Integrating cover crops for weed management in the semiarid U.S. Great Plains: opportunities and challenges. **Weed Science**, v. 68, n. 4, p. 311-323, 2020.
- LEE, Yu-Ting et al. Sulfonylureas Have Antifungal Activity and Are Potent Inhibitors of *Candida albicans* Acetohydroxyacid Synthase. **Journal of Medicinal Chemistry**, v. 56, n. 1, p. 210-219, 2013.
- LEELANANDA, S. P; LINDERT, S. Computational methods in drug discovery. **Beilstein Journal of Organic Chemistry**, v. 12, p. 2694-2718, 2016.
- LIN, L. I-Kuei. A Concordance Correlation Coefficient to Evaluate Reproducibility. **Biometrics**, v. 45, n. 1, p. 255-268, 1989.
- LIU, Y.; LI, Y.; WANG, X. Acetohydroxyacid synthases: evolution, structure, and function. **Applied Microbiology and Biotechnology**, v. 100, n. 20, p. 8633-8649, 2016.
- LIU, Yu-Chao et al. Triazolopyrimidines as a New Herbicidal Lead for Combating Weed Resistance Associated with Acetohydroxyacid Synthase Mutation. **Journal of Agricultural and Food Chemistry**, v. 64, n. 24, p. 4845-4857, 2016.
- LONHIENNE, T. et al. Commercial Herbicides Can Trigger the Oxidative Inactivation of Acetohydroxyacid Synthase. **Angewandte Chemie International Edition**, v. 55, n. 13, p. 4247-4251, 2016.
- MAYEROVÁ, M.; MADARAS, M.; SOUKUP, J. Effect of chemical weed control on crop yields in different crop rotations in a long-term field trial. **Crop Protection**, v. 114, p. 215-222, 2018.
- MITRA, I.; SAHA, A.; ROY, K. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. **Molecular Simulation**, v. 36, n. 13, p. 1067-1079, 2010.

MONTANARI, C. A. Química medicinal: métodos e fundamentos em planejamento de fármacos. São Paulo: EDUSP, 2011. 710 p.

NUNES, C. A.; FREITAS, M. P. Introducing new dimensions in MIA-QSAR: a case for chemokine receptor inhibitors. **European Journal of Medicinal Chemistry**, v. 62, p. 297-300, 2013.

NUNES, C. A. et al. Chemoface: a novel free user-friendly interface for chemometrics. **Journal of the Brazilian Chemical Society**, v. 23, n. 11, p. 2003-2010, 2012.

OLIVEIRA JUNIOR, R. S.; CONSTANTIN, J.; INOUE, M. H. **Biologia e Manejo de Plantas Daninhas**. Capítulo 7 - Mecanismos de ação de herbicidas. 22. ed. Curitiba, Pr: Omnipax Editora, 2011. 348 p.

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

PAGNONCELLI JUNIOR, F. B. et al. Characterization of Ethoxysulfuron Herbicide Selectivity in Common Bean Cultivars. **Planta Daninha**, v. 36, 2018.

QU, Ren-Yu et al. Discovery of New 2-[(4,6-Dimethoxy-1,3,5-triazin-2-yl)oxy]-6-(substituted phenoxy)benzoic Acids as Flexible Inhibitors of Arabidopsis thaliana Acetohydroxyacid Synthase and Its P197L Mutant. **Journal of Agricultural and Food Chemistry**, v. 65, n. 51, p. 11170-11178, 2017.

RAMALHO, T. C. et al. Development of new acetylcholinesterase reactivators: molecular modeling versus in vitro data. **Chemico-Biological Interactions**, v. 185, n. 1, p. 73-77, 2010.

REY-CABALLERO, J. et al. Management Options for Multiple Herbicide-Resistant Corn Poppy (*Papaver rhoeas*) in Spain. **Weed Science**, v. 65, n. 2, p. 295-304, 2017.

ROY, K. et al. Some case studies on application of “ r_m^{2*} ” metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. **Journal of Computational Chemistry**, v. 34, n. 12, p. 1071-1082, 2013.

ROY, K. et al. Be aware of error measures. Further studies on validation of predictive QSAR models. **Chemometrics and Intelligent Laboratory Systems**, v. 152, p. 18-33, 2016.

ROY, K.; KAR, S.; DAS, R. N. A Primer on QSAR/QSPR Modeling: fundamental concepts. **Springerbriefs in Molecular Science**, 2015a. 121 p.

ROY, K.; KAR, S.; DAS, R. N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. **Academic Press**, 2015b. 484 p.

SALES, T. A. **Phospholipases A2 (PLA2) as target enzymes for new anti-inflammatory drugs: a theoretical and experimental study**. 2018. 77 p. Dissertação (Mestrado em Agroquímica) – Universidade Federal de Lavras, Lavras, 2018.

SILVA, T.C. et al. The effect of a chemical additive on the fermentation and aerobic stability of high-moisture corn. **Journal of Dairy Science**, v. 98, n. 12, p. 8904-8912, 2015.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: a new technique for high-accuracy molecular docking. **Journal of Medicinal Chemistry**, v. 49, n. 11, p. 3315-3321, 2006.

TODESCHINI, R. et al. Chemometrics for QSAR Modeling☆. **Comprehensive Chemometrics**, p. 599-634, 2020.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6-7, p. 476-488, 2010.

UEGAKI, R. et al. Effect of Different Silage Storing Conditions on the Oxygen Concentration in the Silo and Fermentation Quality of Rice. **Journal of Agricultural and Food Chemistry**, v. 65, n. 24, p. 4877-4882, 2017.

VAN EVERT, F K et al. Big Data for weed control and crop protection. **Weed Research**, v. 57, n. 4, p. 218-233, 2017.

WATERHOUSE, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. 1, p. W296-W303, 2018.

WEI, W. et al. Design, synthesis and SAR study of novel sulfonylureas containing an alkenyl moiety. **Organic & Biomolecular Chemistry**, Tianjin, v. 14, n. 35, p. 8356-8366, 2016a.

WEI, W. et al. Design, syntheses and biological activities of novel sulfonylureas containing an oxime ether moiety. **Chemical Research in Chinese Universities**, Tianjin, v. 32, n. 2, p. 195-201, 2016b.

WEI, W. et al. Design, synthesis and herbicidal activity study of aryl 2,6-disubstituted sulfonylureas as potent acetohydroxyacid synthase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, Tianjin, v. 27, n. 15, p. 3365-3369, 2017.

WU, Ren-Jun et al. Chemical preparation, biological evaluation and 3D-QSAR of ethoxysulfuron derivatives as novel antifungal agents targeting acetohydroxyacid synthase. **European Journal of Medicinal Chemistry**, v. 162, p. 348-363, 2019.

XAVIER, E. et al. Activity of Antioxidant Enzymes in Euphorbia heterophylla Biotypes and their Relation to Cross Resistance to ALS and Protox Inhibitors. **Planta Daninha**, v. 36, 2018.

ZHOU, S. et al. Research on Controllable Degradation of Novel Sulfonylurea Herbicides in Acidic and Alkaline Soils. **Journal of Agricultural and Food Chemistry**, v. 65, n. 35, p. 7661-7668, 2017.

ZHOU, S. et al. Controllable Soil Degradation Rate of 5-Substituted Sulfonylurea Herbicides as Novel AHAS Inhibitors. **Journal of Agricultural and Food Chemistry**, v. 68, n. 10, p. 3017-3025, 2020.

SEGUNDA PARTE

ARTIGO

**COMPUTER-ASSISTED IMPROVEMENT OF
SULFONYLUREAS WITH ANTIFUNGAL PROPERTIES AND
LIMITED HERBICIDAL ACTIVITY: POTENTIAL
APPLICATION IN FORAGE CONSERVATION**

Adriana Cássia de Faria,^a Joyce K. Daré,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

^a *Departamento de Química, Instituto de Ciências Naturais, Universidade Federal de Lavras, 37200-900, Lavras, MG, Brazil.*

* Corresponding author: matheus@ufla.br

Abstract

This work reports studies at the molecular level of a series of modified sulfonylureas to determine the chemophoric sites responsible for their antifungal and herbicidal activities. For forage conservation, high antifungal potency and low phytotoxicity are required. A molecular modeling study based on multivariate image analysis applied to QSAR (MIA-QSAR) was performed to model these properties, as well as to guide the design of new agrochemical candidates. As a result, the MIA-QSAR models were reliable, robust, and predictive; for antifungal activity, the averages of the main validation parameters were $r^2 = 0.936$, $q^2 = 0.741$, and $r^2_{\text{pred}} = 0.720$, and for herbicidal activity the model was predictive ($r^2_{\text{pred}} = 0.981$ and $r^2_{\text{m}} = 0.944$). From the interpretation of the MIA-plots, 46 novel sulfonylureas with likely improved performance were proposed, from which 9 presented promising calculated selectivity indexes. Docking studies were performed to validate the QSAR predictions and to understand the interaction mode of the proposed ligands with the acetohydroxyacid synthase enzyme.

Keywords: antifungal; herbicide; forage; AHAS; QSAR; docking.

1. INTRODUCTION

Sulfonylureas (Figure 1) comprise a chemical class of substances used in agriculture, especially as herbicides, due to their ability to interfere in the biosynthesis of certain plant amino acids.¹ On the other hand, more recently, these compounds have also been explored as antifungal agents, because of their high inhibitory activity against the acetohydroxyacid synthase (AHAS) enzyme.^{2,3} AHAS is also a key enzyme for the catalysis of the biosynthesis of three essential amino acids in plants, *i.e.*, isoleucine, leucine, and valine.⁴ Its inhibition interrupts the catalytic cycle and protein synthesis in the meristem by blocking cell division, resulting in the inhibition of growth and subsequent death of plants.⁵ The antifungal property provides sulfonylureas the potential of being used as antifungal additives in the silage process. Forage is a vegetal material used for animal feed and, during the silage process, it may be affected by fungi. Therefore, antifungals with limited phytotoxicity are desirable.

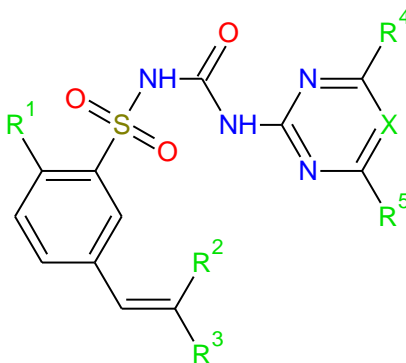


Figure 1. Chemical scaffold of the sulfonylureas studied herein. Changes in R and X affect the antifungal and herbicidal activities of these agrochemicals.

Accordingly, this work reports the rational design of new derivatives of recently synthesized alkenyl-substituted sulfonylureas by Wei et al.,⁶ which aimed for high antifungal activity and, at the same time, limited herbicidal activity. Moreover, a thorough analysis of the molecular features affecting these biological effects and, most importantly, the selectivity index (SI) described as the antifungal/herbicidal activity ratio, is presented herein.

The rational design of bioactive molecules benefits from the use of modeling strategies, such as QSAR (Quantitative Structure-Activity Relationships) and molecular docking, to successfully find potential ligands and biological targets. These techniques are far more assertive than trial and error, not to mention that these approaches are environmentally and economically advantageous. Lastly, they follow a series of rigorous validation protocols widely and exhaustively presented in the literature that guarantee the validity and robustness of the QSAR models, as well as of the possible modes of interaction between ligands and

biological targets proposed by the docking technique.⁷⁻⁹ Regarding the studied data set, although other similar compounds may have available data for antifungal and herbicidal potencies, MIA-QSAR analysis requires highly congeneric compounds and response variables obtained in a unique, systematic manner.

The QSAR method employed herein is based on multivariate image analysis (MIA) descriptors, namely the MIA-QSAR method.¹⁰ MIA descriptors are digital units called pixels, which vary in an image both in orientation (the coordinate in a blackboard) and value (the color, described by the RGB system as a combination of red, green, and blue channels that can numerically range from 0 to 765). Since the image in a MIA-QSAR analysis corresponds to the chemical structure of a molecule, the change in the substituents of this molecule causes modifications in the orientation and value of the pixels in the image, which, in turn, explain the variance in the response variables block (y). Because thousands of pixels usually form an image, a multivariate method for regression is required for modeling purposes, such as partial least squares (PLS) regression.¹¹ A pictographic representation of the MIA-QSAR procedure applied in this work is shown in Figure 2. The atoms in this work were represented as circles with sizes proportional to the respective van der Waals radii and the pixels were numbered proportionally to relevant atomic properties, namely the van der Waals radius (r_{vdW}), electronegativity (ϵ), and the r_{vdW}/ϵ ratio.

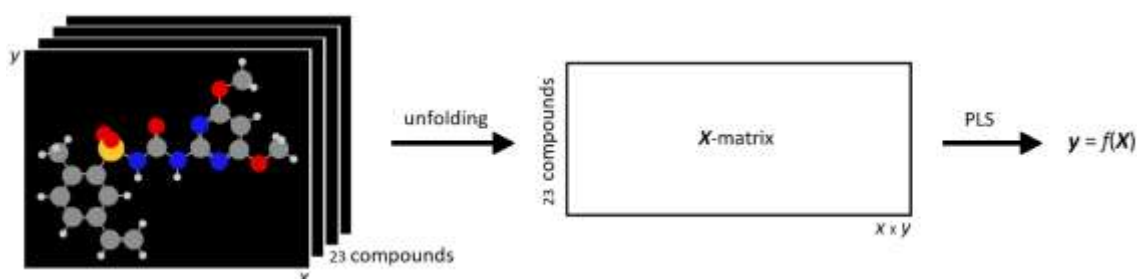


Figure 2. MIA-QSAR workflow from image conversion to a descriptors matrix and PLS regression with y .

It is worth mentioning that a 3D-QSAR study has already been performed to model the antifungal activities of these sulfonylureas against the *Ceratobasidium cornigerum* fungus,⁶ but neither the herbicidal effects nor the selectivity indexes have been computationally evaluated for this class of compounds. In fact, this is the first time that selective agrochemicals are proposed with the main focus on forage conservation.

Finally, but no less importantly, considering that the existence of resistant microorganisms to the available antifungals is a real problem that can affect the efficiency of

sulfonylurea-based additives, the development of new agrochemicals focusing on efficacy and eco-friendship is remarkably necessary,⁵ as it is on forage conservation. Accordingly, molecular contour maps (MIA-plots)^{13,14} describing the molecular moieties most related to the increase or decrease in the observed bioactivities were used to guide the proposition of potential sulfonylurea agrochemicals.

2. MATERIAL AND METHODS

A series of 23 sulfonylureas containing an alkenyl moiety has been previously synthesized and their antifungal activities (expressed in terms of EC₅₀, *i.e.* the concentration for 50% of the maximum effect against *C. cornigerum*, in mg L⁻¹) and herbicidal effects (expressed in terms of % plant growth, measured as fresh weights, at 375 g hm⁻² dosage for pre- and post-emergent application against *Brassica napus* L.) are available in the literature.⁶ The antifungal data were converted into pEC₅₀ (-log EC₅₀) and used in the MIA-QSAR modeling. The herbicidal data were based on the average percentage for pre- and post-emergent values, which were then converted to *I* (% inhibition in decimal scale), *pI* (-log *I*), and probit unit.¹⁵ The best modeling performance was found with the *I* data, which were further used to obtain the selectivity indexes (SI = pEC₅₀/*I*). The compound data set and respective biological values used in the MIA-QSAR analyses are given in Table 1. For the MIA-QSAR modeling, the following procedure was carried out: the molecular images were designed using the GaussView program¹⁶ and saved as bitmaps in a 200 × 228 pixels blackboard. These images should be congruently aligned to make substituents the only variable moieties that explain the variance in the *y* block (the antifungal and herbicidal activities column vectors) (Figure 3). The atoms in the molecules were represented as circles sized proportionally to their respective van der Waals radii (*r*_{vdW}), and the pixels for each atom type were numbered to encode the *r*_{vdW}, the Pauling's electronegativity (ϵ), and the *r*_{vdW}/ ϵ ratio. For example, the hydrogen atom was represented as a small circle numbered with 120 (*r*_{vdW} for H = 120 pm), 220 (ϵ _H = 2.2), and 545 (*r*_{vdW}/ ϵ for H = 54.5). The 23 × 200 × 228 three-way array was unfolded to a 23 × 45,600 matrix (**X**), which was split into training and test sets, according to the splitting previously reported in the literature.⁶ The training set matrix was regressed against the *y* blocks using PLS regression and the calibration parameters were then used for prediction purposes. It is worth mentioning that three MIA-QSAR models were built for each dependent variable, considering the descriptor matrices that encode *r*_{vdW}, ϵ , and *r*_{vdW}/ ϵ . An applicability domain test using Student's residuals and leverage analyses was performed to assess the presence of possible outliers in the data set.¹⁷ A casualty test was also

performed to guarantee that the models were not overfitted: the y block was randomized and regressed against the unmodified X matrix. The correlation obtained from the plot of experimental *versus* predicted values should be significantly poorer than that of the real calibration, which is evaluated by the ${}^c r_p^2$ parameter (values above 0.5 are acceptable), according to Equation 1.¹⁸

$${}^c r_p^2 = r \times (r^2 - r_{y\text{-rand}}^2)^{1/2} \quad (1)$$

where r corresponds to the correlation coefficient for the plot of experimental *versus* predicted values, r^2 and $r_{y\text{-rand}}^2$ stand for the determination coefficients obtained from the corresponding plots considering the intact and randomized y block, respectively.

Table 1. The series of 23 sulfonylureas with antifungal activity against *C. cornigerum* (pEC_{50}), herbicidal activity against *B. napus* (I), and selectivity index (SI).

Compound	R ¹	R ²	R ³	R ⁴	R ⁵	X	pEC_{50}	I	SI
1	CH ₃	H	H	OCH ₃	OCH ₃	C-H	1.655	0.811	2.04
2	CH ₃	H	CN	OCH ₃	OCH ₃	C-H	1.644	0.776	2.12
3	CH ₃	H	CN	CH ₃	CH ₃	C-H	1.596	0.638	2.50
4 ^a	CH ₃	H	COOCH ₃	OCH ₃	OCH ₃	C-H	1.612	0.453	3.56
5	CH ₃	H	COOCH ₃	CH ₃	CH ₃	C-H	1.688	0.230	7.34
6	CH ₃	H	CH ₃	OCH ₃	OCH ₃	C-H	1.644	0.917	1.79
7	CH ₃	H	H	CH ₃	CH ₃	C-H	1.758	0.308	5.72
8	CH ₃	H	H	H	CH ₃	C-H	1.775	0.329	5.40
9 ^a	CH ₃	H	CH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1.651	0.711	2.32
10 ^a	CH ₃	H	CH ₂ CH ₃	CH ₃	CH ₃	C-H	1.865	0.477	3.91
11	CH ₃	H	CH ₂ CH ₃	H	CH ₃	C-H	1.877	0.416	4.52
12	CH ₃	H	COOCH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1.743	0.418	4.17
13	CH ₃	H	COOCH ₂ CH ₃	CH ₃	CH ₃	C-H	1.818	0.167	10.89
14 ^a	CH ₃	H	COOCH ₂ CH ₃	H	CH ₃	C-H	1.830	0.106	17.35
15	CH ₃	H	CH ₂ CH ₂ CH ₃	OCH ₃	OCH ₃	C-H	1.751	0.279	6.28
16	CH ₃	H	CH ₂ CH ₂ CH ₃	CH ₃	CH ₃	C-H	1.903	0.225	8.46
17	CH ₃	H	C ₆ H ₅	OCH ₃	OCH ₃	C-H	1.860	0.514	3.62
18	CH ₃	CH ₃	COOCH ₃	OCH ₃	OCH ₃	C-H	1.063	0.525	2.03
19	CH ₃	CH ₃	COOCH ₃	CH ₃	CH ₃	C-H	1.154	0.318	3.63
20	Cl	H	H	OCH ₃	OCH ₃	C-H	1.763	0.975	1.81
21	Cl	H	CH ₃	OCH ₃	OCH ₃	C-H	1.592	0.936	1.70

22	Cl	H	H	CH ₃	OCH ₃	N	1.428	0.925	1.54
23	Cl	H	H	CH ₃	OCH ₃	C-Br	1.756	0.075	23.41

^a Test set compounds (the same used in ref. 6).

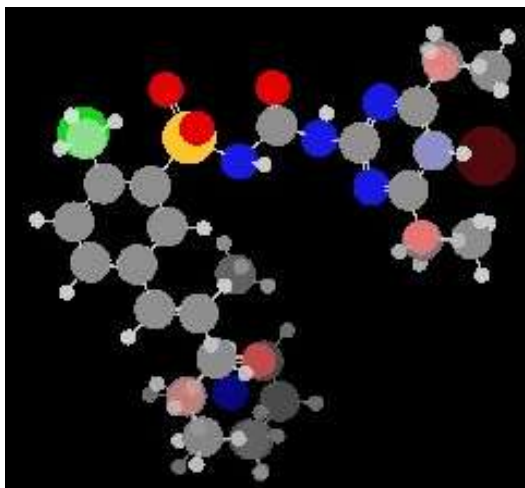


Figure 3. Superimposed images of the 23 sulfonyleureas studied in this work. Note that the coincident substructure of the compounds is congruently aligned.

The models were internally (leave-one-out cross-validation – LOOCV) and externally (using the test set of compounds and a bootstrapping procedure) validated. The determination coefficient q^2 in the LOOCV is generally acceptable if larger than 0.5,¹⁹ as well as the r^2_{pred} obtained in the external validation. The r^2_{pred} should be close to the optimum correlation coefficient, which is obtained forcing the regression line to cross the origin point (r^2_0); this can be evaluated through the r^2_m parameter (values above 0.5 are acceptable), according to Equation 2.²⁰

$$r^2_m = r^2_{\text{pred}} \times [1 - (r^2_{\text{pred}} - r^2_0)^{1/2}] \quad (2)$$

where r^2_{pred} and r^2_0 correspond to the determination coefficients for the plot of experimental *versus* predicted values with lines unforced and forced to pass through the origin, respectively.

To gain insight into the robustness of the models, additional statistical parameters for external validation were evaluated, namely the mean absolute error (MAE),²¹ the slopes k and k' of the curves of experimental *vs.* predicted values (these parameters should range from 0.85 to 1.15),²² Q^2_{Fn} (corresponding to different metrics that account for the difference between experimental and predicted data, which are considered acceptable if higher than 0.6),^{23,24} and the concordance correlation coefficient (CCC, which verifies how small the differences are

between experimental data and external data set predictions, independently of their range, and should be higher than 0.85).²³⁻²⁵

The stability of the models was also evaluated by a bootstrapping procedure, in which multiple test sets were randomly selected (20 times), and the calibration and validation steps were subsequently performed for each cycle (mean statistical values were collected). This procedure is particularly necessary for small data sets, to avoid the risk of atypical behavior, *e.g.*, due to chance correlation, lack of robustness to resampling and cross-validation.²⁶

The chemical features affecting the response variables were analyzed using the MIA-plots.^{13,14} The **b**-plot consists of a contour map describing the PLS regression coefficients, *i.e.*, it denotes the chemical space influencing either negatively or positively the response variable. In turn, the VIP-plot indicates the molecular moieties mostly affecting the response variable. The entire modeling procedure was performed using the Chemoface program.²⁶

A molecular docking study was carried out to explain the binding affinity of the sulfonylureas to the AHAS (EC 2.2.1.6). The enzyme has three cofactors, namely the thiamine diphosphate (ThDP), Mg²⁺, and FAD. The crystal structure of *Saccharomyces cerevisiae* acetohydroxyacid synthase (*Sc*AHAS) in complex with sulfometuron-methyl, at a resolution of 2.34 Å, was obtained from PDB (code: 1T9C).²⁷ We used this complex because of the structural similarity of sulfometuron-methyl with the studied compounds. The crystal structure of *B. napus* AHAS (*Bn*AHAS) was obtained by homology modeling.²⁸ The search space of the simulation in the docking studies was defined as a region of 12.0 Å around the sulfonylurea herbicides. The cavity prediction algorithm (MolDock) identified the potential binding sites with a volume of 215.04 Å³ and 114.18 Å³, in the *Sc*AHAS and *Bn*AHAS, respectively. Acid/basic enzyme residues were considered at neutral pH; for histidines, we considered the *N*-delta nitrogen protonated. During simulation for pose generation, 1500 maximum iterations were used by selecting a population size of 50 and the poses generated were added to the population if the energy value was below the threshold of 100.0. The SE parameters were set at 300 steps with a neighbor distance factor of 1.0. The modeling started with a systematic search in the Swiss Bioinformatics Resource Portal (Expasy) to obtain the primary structure of the *Bn*AHAS using the UniProt database²⁹ as a search tool. This profile was then used to search the PDB for known protein structures using the target sequence as the query. The primary sequence of the target protein was aligned with the template protein using the SIM similarity program.³⁰ The templates were ranked based on the alignment score and the structural quality according to WHAT_CHECK.³¹ The *Bn*AHAS was constructed using SWISS-MODEL server.³² The structures of the ligands of Table 1 were prepared and

optimized with the AMBER force field (GAFF) parameters and AM1-BCC charges. Subsequently, the ligands were docked into the AHAS binding site using the Molegro Virtual Docker software,³³ which is designed for the prediction of the most likely binding conformation of a ligand towards a macromolecule.³⁴ MolDock Score [GRID] function is a hybrid search algorithm that combines the differential evolution optimization technique with a cavity prediction algorithm during the searching procedure, which allows fast and accurate recognition of potential binding modes, the poses. It is derived from the Piecewise Linear Potential (PLP), a simplified potential whose parameters are fit to protein-ligand structures and binding data scoring functions, and further extended in the GEMDOCK program (Generic Evolutionary Method for molecular DOCK) with a new hydrogen bonding term and new charge schemes.³³ The grid-based scoring functions pre-calculate potential-energy values on an evenly spaced cubic grid to speed up calculations. The energy potential was evaluated using tri-linear interpolation among relevant grid points. A grid of 0.30 Å in resolution and 12 Å in radius from the inhibitor and the binding site was used. MolDock SE was used as a searching algorithm, and the maximum number of poses to generate was 100.

3. RESULTS AND DISCUSSION

The first step before performing the MIA-QSAR modeling was to verify whether the series of compounds fell within a homogeneous sample space. Accordingly, an applicability domain test using sample leverages and studentized residuals was performed (Supporting Information), and compound **23** exhibited an anomalous behavior considering the herbicidal data. However, since this is a key compound, as it is the only one bearing an X = C□Br and it possesses the lowest herbicidal activity, it was retained in the model. Therefore, the modeling of the antifungal (pEC₅₀) and herbicidal (*I*) activities proceeded with the unmodified data set presented in Table 1. It is worth mentioning that although other organisms have been studied by Wei *et al.*,⁶ only the data for *C. cornigerum* and *B. napus* L. could be properly modeled.

To evaluate the modeling capability of the MIA descriptors, the pEC₅₀ and *I* values for the 23 compounds were regressed against the independent variables, yielding satisfactory statistical parameters (Figure 4). It is worth mentioning that the three MIA descriptor types (r_{vdW} , ϵ , and r_{vdW}/ϵ) demonstrated similar modeling performance, indicating that both hydrophobic and electrostatic interactions in the enzyme pocket are related to the antifungal and herbicidal activities. To gain further insight into the predictive ability of the MIA-QSAR models, the data set was split into training and test subsets in the same way as in the literature, for comparison purposes.⁶ The most relevant parameters to take into consideration when

evaluating QSAR models were found to be meaningful in the modeling of the pEC₅₀ data (average for the three models: $r^2 = 0.936$, $q^2 = 0.741$, and $r^2_{\text{pred}} = 0.720$ using three or four PLS components), while the corresponding results using the CoMSIA method were $r^2 = 0.989$, $q^2 = 0.585$, and $r^2_{\text{pred}} = 0.856$ for 6 PLS components.⁶ It is natural that q^2 and r^2_{pred} are smaller than r^2 , since the samples left out are not included in the calibration step. The other parameters useful to attest the modeling and prediction performance were also, in general, acceptable, with a mean pScore (average of the statistical parameters with a cut-off value of 0.5, in percent) of 66.6% (Table 2). Notably, the MIA-QSAR models for the herbicidal activities were quite predictive, with an average r^2_{pred} of 0.981 and r^2_m of 0.944 (refer to Table 2 for the complete statistical data and to Supporting Information for the predicted antifungal and herbicidal activities).

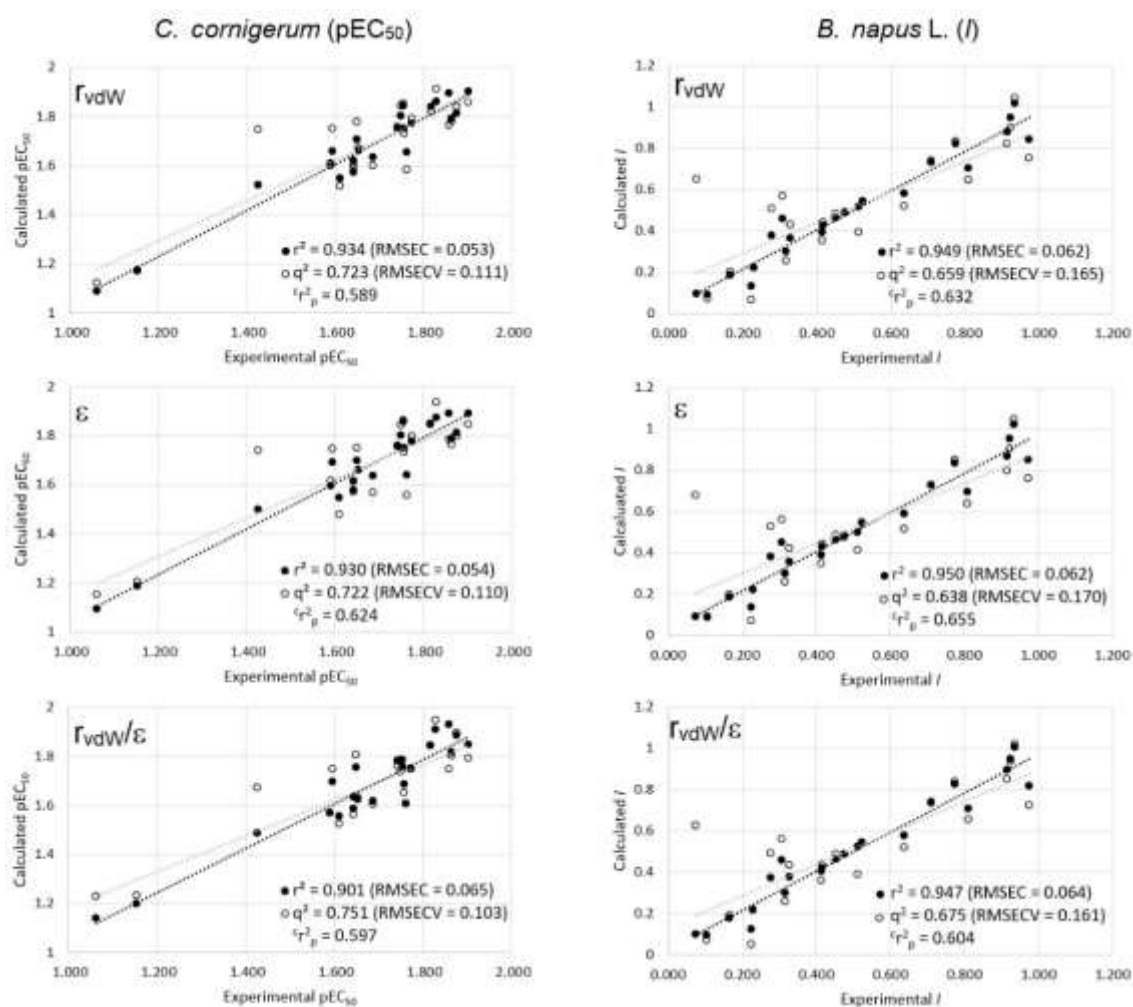


Figure 4. Correlation plots of experimental \times calculated antifungal and herbicidal activities for the 23 sulfonylurea derivatives using MIA descriptors based on r_{vdW} , ϵ , and r_{vdW}/ϵ .

Table 2. Statistical results of the MIA-QSAR models obtained using the four test set compounds of the literature (**4**, **9**, **10**, and **14**).

Parameters	<i>C. cornigerum</i>			<i>B. napus L.</i>		
	r_{vdW}	ε	r_{vdW}/ε	r_{vdW}	ε	r_{vdW}/ε
PLS components	4	4	3	5	5	3
RMSEC	0.052	0.054	0.059	0.068	0.068	0.081
r^2	0.942	0.940	0.926	0.944	0.944	0.919
RMSE _{y-rand}	0.128	0.133	0.142	0.174	0.156	0.201
r^2_{y-rand}	0.638	0.620	0.559	0.619	0.689	0.504
$^c r^2_p$	0.536	0.548	0.583	0.554	0.491	0.618
RMSECV	0.118	0.118	0.101	0.188	0.195	0.179
q^2	0.714	0.722	0.787	0.593	0.569	0.611
RMSEP	0.079	0.081	0.083	0.026	0.019	0.055
r^2_{pred}	0.731	0.717	0.711	0.997	0.994	0.953
r^2_m	0.648	0.641	0.655	0.936	0.976	0.921
CCC	0.816	0.804	0.797	0.994	0.996	0.971
MAE	0.012	0.018	0.024	0.008	0.004	0.012
k	1.004	1.008	0.984	0.966	0.985	0.955
k'	0.994	0.990	1.014	1.034	1.014	1.036
Q^2_{F1}	0.673	0.650	0.637	0.988	0.993	0.944
Q^2_{F2}	0.485	0.448	0.428	0.986	0.992	0.937
Q^2_{F3}	0.870	0.861	0.856	0.992	0.996	0.964
pScore (%) ^a	65.7	65.7	68.4	77.0	75.8	77.6

^a Average of $^c r^2_p$, q^2 , r^2_{pred} , and r^2_m , in percent.

Since the sulfonylureas data set comprises a small number of samples, the lack of only a few compounds in the training subset may considerably disturb the model. Accordingly, to assess the model robustness, as well as to attest that the four test set samples used earlier did not yield good external validation outputs due to chance correlation, a bootstrapping procedure was employed by randomly selecting four test samples twenty times; the *sample function*, available in R program³⁵ was used for this purpose. The resulting models demonstrated acceptable average performance in most of the internal and external validation parameters, except for r^2_m and CCC, but with standard deviations in general higher than 20% (Table 3). This instability indicates that the model quality is sensitive to changes in the

compounds that compose the test set, *i.e.*, each compound in the data set contributes importantly to the MIA-QSAR model. Therefore, any prediction for congeneric agrochemical candidates should be carried out using the models built with all 23 molecules included.

Table 3. Statistical results (mean of 20 cycles) of the MIA-QSAR models obtained through a bootstrapping procedure with four randomly selected test set samples.

Parameters	<i>C. cornigerum</i>			<i>B. napus L.</i>		
	r_{vdW}	ε	r_{vdW}/ε	r_{vdW}	ε	r_{vdW}/ε
PLS components	4.8 ± 1.1	4.3 ± 1.3	4.5 ± 1.0	4.0 ± 0.6	4.3 ± 0.8	3.7 ± 0.7
RMSEC	0.04 ± 0.01	0.05 ± 0.02	0.04 ± 0.01	0.06 ± 0.01	0.06 ± 0.01	0.07 ± 0.01
r^2	0.95 ± 0.03	0.94 ± 0.04	0.95 ± 0.03	0.95 ± 0.02	0.95 ± 0.02	0.94 ± 0.03
$RMSE_{y\text{-rand}}$	0.12 ± 0.02	0.12 ± 0.02	0.12 ± 0.02	0.18 ± 0.01	0.17 ± 0.01	0.18 ± 0.01
$r^2_{y\text{-rand}}$	0.61 ± 0.07	0.61 ± 0.08	0.63 ± 0.07	0.56 ± 0.03	0.58 ± 0.04	0.54 ± 0.04
${}^c r^2_p$	0.65 ± 0.05	0.55 ± 0.06	0.55 ± 0.06	0.60 ± 0.03	0.58 ± 0.03	0.61 ± 0.03
RMSECV	0.12 ± 0.04	0.12 ± 0.03	0.11 ± 0.04	0.18 ± 0.02	0.18 ± 0.02	0.17 ± 0.02
q^2	0.62 ± 0.28	0.60 ± 0.26	0.66 ± 0.28	0.59 ± 0.09	0.57 ± 0.11	0.61 ± 0.09
RMSEP	0.14 ± 0.08	0.13 ± 0.07	0.14 ± 0.08	0.15 ± 0.08	0.16 ± 0.09	0.16 ± 0.08
r^2_{pred}	0.50 ± 0.38	0.62 ± 0.33	0.45 ± 0.39	0.65 ± 0.34	0.65 ± 0.33	0.64 ± 0.33
r^2_m	0.27 ± 0.35	0.33 ± 0.39	0.28 ± 0.38	0.48 ± 0.36	0.47 ± 0.35	0.47 ± 0.36
CCC	0.38 ± 0.43	0.43 ± 0.44	0.41 ± 0.37	0.69 ± 0.33	0.68 ± 0.33	0.69 ± 0.31
MAE	0.08 ± 0.07	0.07 ± 0.06	0.08 ± 0.07	0.06 ± 0.04	0.05 ± 0.04	0.06 ± 0.03
k	0.98 ± 0.06	0.98 ± 0.05	0.98 ± 0.06	0.95 ± 0.17	0.95 ± 0.18	0.96 ± 0.18
k'	1.02 ± 0.06	1.01 ± 0.05	1.02 ± 0.06	0.96 ± 0.07	0.95 ± 0.08	0.94 ± 0.08
Q^2_{F1}	0.63 ± 0.27	0.62 ± 0.27	0.62 ± 0.28	0.71 ± 0.22	0.71 ± 0.22	0.71 ± 0.21
Q^2_{F2}	0.65 ± 0.33	0.56 ± 0.32	0.54 ± 0.38	0.73 ± 0.22	0.71 ± 0.22	0.70 ± 0.22
Q^2_{F3}	0.57 ± 0.32	0.63 ± 0.28	0.61 ± 0.30	0.77 ± 0.17	0.74 ± 0.24	0.73 ± 0.22
pScore (%) ^a	50.7	52.6	48.3	58.1	56.7	58.4

^a Average of ${}^c r^2_p$, q^2 , r^2_{pred} , and r^2_m , in percent.

Since the models have been vigorously validated, they are ready to predict the pEC₅₀ and *I* data of congeneric proposals. To guide the design of new agrochemical candidates, the chemical features influencing both the antifungal and herbicidal activities were analyzed through the MIA-plots (Figure 5). The variable importance in projection (VIP) scores indicate the molecular moieties that most explain the variance in the activities, whereas the PLS regression coefficients (**b**) identify the substituents contributing to increasing or decreasing the activities. Since the MIA-plots based on r_{vdW} , ε , and r_{vdW}/ε presented a similar profile,

only the former is shown in Figure 5 (refer to the Supporting Information for all MIA-plots). Regarding the antifungal activity, it is remarkable from the VIP plot that a methyl substituent at R^2 plays a major role in describing the pEC_{50} and, according to the **b** plot, this influence is negative (it contributes to decreasing pEC_{50}). On the other hand, the substituents COOEt, Et, and Pr at R^3 play a secondary role in describing the pEC_{50} data, but their contributions are in favor of the antifungal activity. Considering the herbicidal activity, which is desirable to be small, the substituents that mostly explain the variance in the y -block are Pr at R^3 (decreasing I), Me (decreasing I), and OMe (increasing I) at R^5 . In addition, $R^1 = Cl$ and $X = C-Br$ contribute to increase and decrease I , respectively. Based on these findings, as well as by combining the substructures of the best antifungals ($pEC_{50} > 1.8$) and worst herbicides ($I < 0.3$) of Table 1, 46 unprecedented agrochemical derivatives were proposed and the respective pEC_{50} and I values were predicted using the MIA-QSAR models (Table 4).

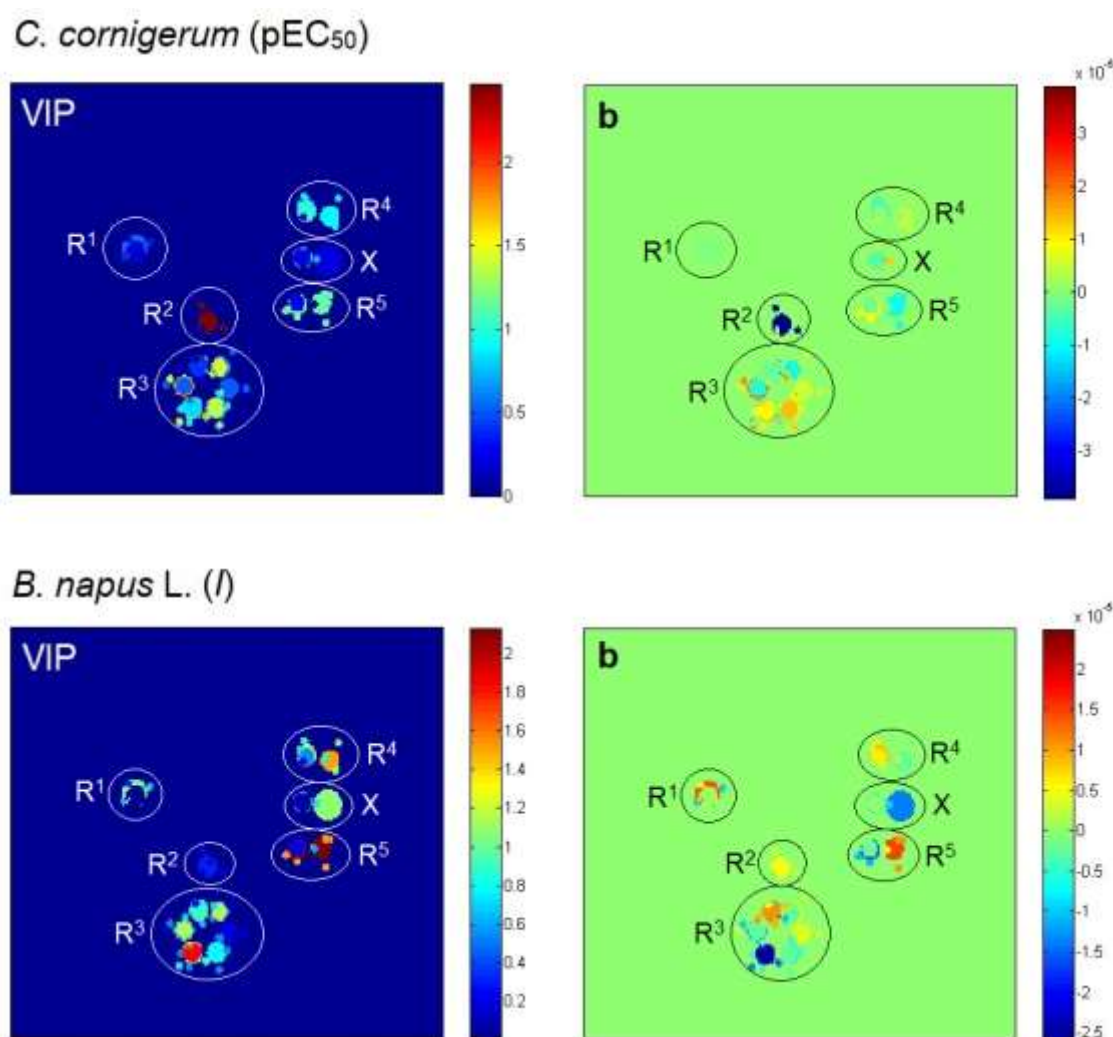


Figure 5. MIA-plots indicating how and how much the sulfonylurea substituents affect the antifungal and herbicidal activities.

Table 4. Proposed sulfonylurea candidates based on the MIA-plot outcomes and the combination of substructures of the most suitable compounds previously synthesized.^a The calculated data correspond to mean values obtained from the three MIA-QSAR models, ranked in descending order of selectivity.

Proposal	R ¹	R ²	R ³	R ⁴	R ⁵	X	pEC ₅₀	I	SI
P1	CH ₃	H	COOCH ₂ CH ₃	H	CH ₃	C-Br	1.926 ± 0.021	-0.323 ± 0.039	> 50
P2	CH ₃	H	COOCH ₂ CH ₃	CH ₃	CH ₃	C-Br	1.869 ± 0.014	-0.232 ± 0.033	> 50
P3	Cl	H	COOCH ₂ CH ₃	H	CH ₃	C-Br	1.903 ± 0.016	-0.189 ± 0.018	> 50
P4	Cl	H	CH ₂ CH ₂ CH ₃	CH ₃	CH ₃	C-Br	1.858 ± 0.040	-0.148 ± 0.011	> 50
P5	Cl	H	COOCH ₂ CH ₃	CH ₃	CH ₃	C-Br	1.846 ± 0.014	-0.097 ± 0.012	> 50
P6	CH ₃	H	CH ₂ CH ₃	H	CH ₃	C-Br	1.887 ± 0.038	-0.018 ± 0.045	> 50
P7	CH ₃	H	CH ₂ CH ₂ CH ₃	H	CH ₃	C-H	1.913 ± 0.002	0.038 ± 0.004	50.34
P8	CH ₃	H	H	CH ₃	CH ₃	C-Br	1.733 ± 0.046	0.044 ± 0.040	39.39
P9	CH ₃	H	CH ₂ CH ₃	CH ₃	CH ₃	C-Br	1.829 ± 0.010	0.073 ± 0.039	25.05
P10	Cl	H	H	H	CH ₃	C-Br	1.766 ± 0.023	0.087 ± 0.026	20.30
P11	CH ₃	H	C ₆ H ₅	OCH ₃	OCH ₃	C-Br	1.939 ± 0.013	0.099 ± 0.048	19.59
P12	CH ₃	H	COOCH ₂ CH ₃	CH ₃	OCH ₃	C-Br	1.718 ± 0.016	0.099 ± 0.052	17.35
P13	Cl	H	CH ₂ CH ₃	H	CH ₃	C-Br	1.864 ± 0.033	0.117 ± 0.025	15.93
P14	CH ₃	H	C ₆ H ₅	H	CH ₃	C-H	2.044 ± 0.033	0.174 ± 0.014	11.75
P15	CH ₃	H	C ₆ H ₅	CH ₃	OCH ₃	C-Br	1.861 ± 0.023	0.185 ± 0.064	10.06
P16	Cl	H	H	CH ₃	CH ₃	C-Br	1.710 ± 0.047	0.179 ± 0.020	9.55
P17	Cl	H	CH ₂ CH ₂ CH ₃	CH ₃	OCH ₃	C-Br	1.704 ± 0.016	0.187 ± 0.030	9.11
P18	Cl	H	CH ₂ CH ₃	CH ₃	CH ₃	C-Br	1.807 ± 0.005	0.208 ± 0.018	8.69
P19	Cl	H	C ₆ H ₅	OCH ₃	OCH ₃	C-Br	1.917 ± 0.009	0.233 ± 0.026	8.23
P20	Cl	H	COOCH ₂ CH ₃	CH ₃	OCH ₃	C-Br	1.695 ± 0.010	0.234 ± 0.031	7.24
P21	Cl	H	CH ₂ CH ₂ CH ₃	CH ₃	CH ₃	C-H	1.833 ± 0.033	0.263 ± 0.029	6.97
P22	Cl	H	COOCH ₂ CH ₃	CH ₃	CH ₃	C-H	1.821 ± 0.009	0.314 ± 0.026	5.80
P23	Cl	H	C ₆ H ₅	CH ₃	OCH ₃	C-Br	1.838 ± 0.018	0.320 ± 0.043	5.74
P24	CH ₃	H	CH ₂ CH ₂ CH ₃	H	OCH ₃	C-H	1.761 ± 0.024	0.368 ± 0.019	4.79
P25	CH ₃	H	H	CH ₃	OCH ₃	C-Br	1.581 ± 0.025	0.375 ± 0.060	4.22
P26	CH ₃	H	COOCH ₂ CH ₃	H	OCH ₃	C-H	1.749 ± 0.049	0.419 ± 0.021	4.17
P27	CH ₃	H	CH ₂ CH ₃	CH ₃	OCH ₃	C-Br	1.676 ± 0.033	0.408 ± 0.059	4.11
P28	CH ₃	H	C ₆ H ₅	H	OCH ₃	C-H	1.892 ± 0.057	0.505 ± 0.033	3.75

P29	CH ₃	H	CH ₂ CH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.704 ± 0.006	0.460 ± 0.012	3.70
P30	Cl	H	H	H	CH ₃	C–H	1.741 ± 0.012	0.499 ± 0.013	3.49
P31	Cl	H	CH ₂ CH ₃	H	CH ₃	C–H	1.838 ± 0.041	0.529 ± 0.014	3.47
P32	CH ₃	H	COOCH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.692 ± 0.021	0.511 ± 0.014	3.31
P33	CH ₃	H	C ₆ H ₅	CH ₃	OCH ₃	C–H	1.836 ± 0.029	0.597 ± 0.027	3.08
P34	Cl	H	CH ₂ CH ₃	CH ₃	OCH ₃	C–Br	1.652 ± 0.029	0.543 ± 0.037	3.04
P35	Cl	H	C ₆ H ₅	OCH ₃	OCH ₃	C–H	1.891 ± 0.020	0.646 ± 0.012	2.93
P36	Cl	H	CH ₂ CH ₃	CH ₃	CH ₃	C–H	1.781 ± 0.016	0.620 ± 0.020	2.87
P37	Cl	H	CH ₂ CH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.679 ± 0.012	0.598 ± 0.011	2.81
P38	Cl	H	COOCH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.669 ± 0.018	0.646 ± 0.009	2.58
P39	Cl	H	C ₆ H ₅	CH ₃	OCH ₃	C–H	1.813 ± 0.026	0.732 ± 0.006	2.48
P40	CH ₃	H	CH ₂ CH ₃	H	OCH ₃	C–H	1.710 ± 0.068	0.725 ± 0.027	2.36
P41	CH ₃	H	H	H	OCH ₃	C–H	1.612 ± 0.019	0.695 ± 0.028	2.32
P42	CH ₃	H	CH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.653 ± 0.040	0.816 ± 0.021	2.03
P43	CH ₃	H	H	CH ₃	OCH ₃	C–H	1.555 ± 0.014	0.786 ± 0.022	1.98
P44	Cl	H	H	H	OCH ₃	C–H	1.590 ± 0.013	0.830 ± 0.008	1.92
P45	Cl	H	CH ₂ CH ₃	CH ₃	OCH ₃	C–H	1.627 ± 0.037	0.955 ± 0.002	1.70
P46	Cl	H	H	CH ₃	OCH ₃	C–H	1.533 ± 0.015	0.922 ± 0.002	1.66

^a Suitable compounds for substructural combination, which yields **P1–P46**: pEC₅₀ > 1.8 (**10**, **11**, **13**, **14**, **16**, **17**); *I* < 0.30 (**8**, **13**, **14**, **15**, **16**, **23**); pEC₅₀ > 1.8 and *I* < 0.30 (**13**, **14**, **16**).

Considering that desirable agrochemical candidates for use in forage conservation should be highly selective (large pEC₅₀ and small *I*), the proposed sulfonylurea derivatives **P1–P9** are expected to exhibit better performance than compound **23** (SI = 23.41), which is the best experimentally known derivative according to this criterion. The negative SI values in Table 4 are due to the negative modeled *I*, which should be rather interpreted as “too low *I*”. From these proposals, **P1**, **P3**, and **P7** are expected to perform even better than compound **16** (the best antifungal in the series) and **23** (the worst herbicide in the series). Accordingly, at least three synthetic targets stand out as promising antifungals with limited herbicidal activity. The rationale behind this behavior can be interpreted through docking studies, which can also provide further support to the QSAR outcomes.

To obtain reliable docking outcomes, the primary sequences of *B. napus* and *Arabidopsis thaliana* AHAS were aligned and showed 89.0% identity (Supporting

Information). Therefore, considering that aligned sequences over 40 – 50% are nearly identical, the alignments were then considered accurate in a structural sense.³⁶ The SWISS-MODEL server was used to generate the crystal structure of *Bn*AHAS, while the 3EA4 structure from PDB was used for modeling.³⁷ The quality of the homology model was evaluated by different tools, such as the Ramachandran plot. The modeled protein represents 88.4% of overall residues in the most favored regions, and 12.9% in allowed regions, representing a good quality model (Figure 6). The RMSD value of 1.67 Å was found in the best superimposition for protein structures; lower values (≤ 2 Å) mean better alignment between the structures. After the validation of the stereochemical quality of the protein structure, a comparative analysis between the AHAS active sites of the plant and fungus was performed. The herbicide-binding site is located at the interface of two adjoining catalytic subunits at the base of a hydrophobic tunnel and the sulfonylurea herbicides bind at its entrance. Sulfonylurea herbicides bind next to 13 amino acids residues: Gly116, Val191, Pro192, Ala195, Ala200, Phe 201, Ala251, Met354, Asp379, Arg380, Met582, Val583, and Trp586. The herbicide-binding site residues are highly conserved; however, the Ile651, Pro652, and Ser653 amino acids of the *Bn*AHAS, located in the C-terminal arm next to the active site, are substituted by Val655, Ala656, and Gly657 in the *Sc*AHAS.³⁸ These amino acids might be important for understanding the molecular origin of the selectivity differences of the studied compounds.

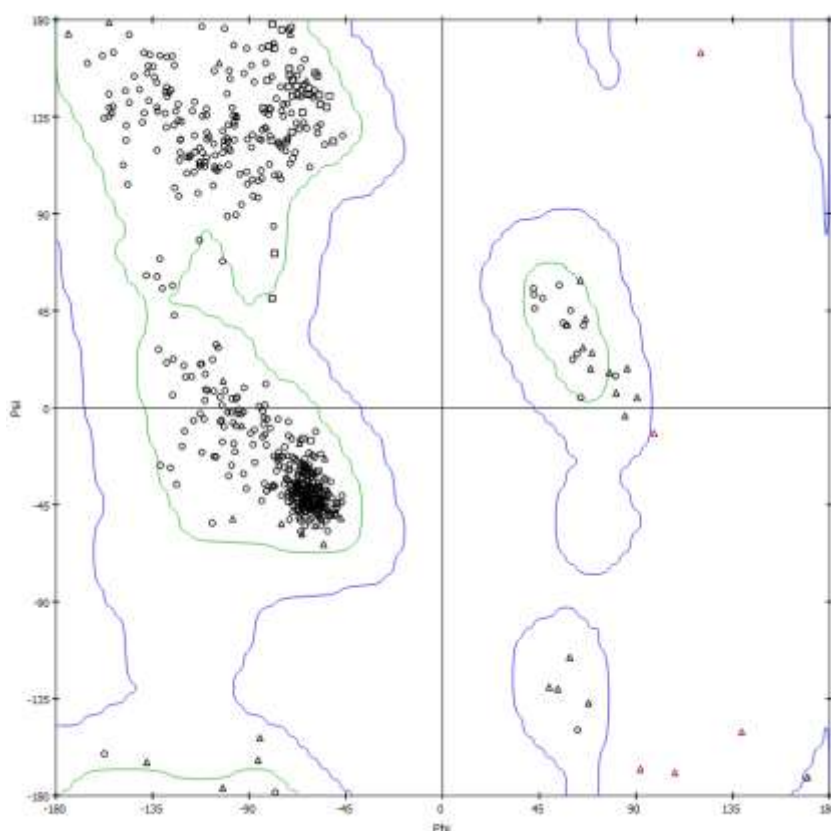


Figure 6. Ramachandran plot demonstrating the quality of the homology modeling for *BnAHAS*.

The docking studies were validated by redocking the sulfometuron-methyl into the herbicide-binding site and superimposing it onto the co-crystallized complex. Two conformations were considered: the original from PDB (SM1) and the conformation after minimization (SM2). Programs capable of returning poses below a pre-selected RMSD value from the known conformation (usually $< 2 \text{ \AA}$ depending on ligand size) are considered successful in their performance. The poses (SM1 and SM2) were obtained, and the complex structures were superimposed to the crystal structure. RMSD values of 0.36 \AA for SM1, and 0.21 \AA for SM2-crystal were observed. These RMSD values demonstrate MVD as accurate in reproducing the experimental binding mode. Therefore, the same method was used in the docking process. The poses were selected according to three criteria: 1) the orientation of the monosulfuron-ester and sulfometuron-methyl into the herbicide-binding site (experimental information); 2) the pose with the lowest energy; and 3) the pose with similar spatial position for all compounds. All compounds occupied a similar location in the herbicide-binding site as for the monosulfuron-ester and bensulfuron methyl when bound to crystal structure 3EA4 and 5FEM, respectively. The intermolecular interactions between the most promising ligands **P1**,

P3, and **P7**, and the two proteins include hydrogen bonds, van der Waals contacts, and electrostatic interactions (Figure 7): i) all compounds form π - π stacking interactions between pyrimidine and Trp586 in *ScAHAS*, and Trp574 in *BnAHAS*; ii) all compounds form hydrogen bond with Arg380A in *ScAHAS*, and Arg377A in *BnAHAS*; iii) compounds with Me and OMe groups at R⁴ and/or R⁵ positions form alkyl interactions with Val583A and Met354A in *ScAHAS*, and Val571A and Met351A in *BnAHAS*; iv) compounds with Br and non-hydrogen atoms at R²/R³ position form unfavorable bump interaction with Met582A in *ScAHAS*, and Met570A in *BnAHAS*; v) all compounds form hydrogen bond with Ser653A in *ScAHAS*; vi) compounds with bulky groups at R²/R³ position interact with Phe578A and Tyr579A in *ScAHAS*, and Phe590A and Tyr591A in *BnAHAS*; iv) all compounds form π -alkyl interactions between the phenyl ring and any amino acid residue of chain B. The interactions between compounds **16**, **23**, **P1**, **P3**, and **P7**, and the herbicide-binding site residues are shown in Figure 6. The alkenyl moiety is important for antifungal activity, but it suppresses the herbicidal potency. This appears to be related to subtle differences in the surrounding environment, *i.e.*, the alkenyl moiety is located next to Val191B, Pro192B, Ala195, Ala200, and Phe201B in *ScAHAS*, but it is close to Val196, Pro197B, Agr199B, Met200B, Ala205B, and Asp376 in *BnAHAS*. The output of the docking programs produces one (or several) different poses for every ligand, and further ranks differentiate compounds based on their scoring functions to provide a fast and crude estimation of the binding affinity. Comparing the energy values obtained, we found that most of the *ScAHAS*-ligand complexes are more stable than *BnAHAS*-ligand complexes, which explains the selectivity indexes expected for these compounds (Figure S3, Supporting Information).

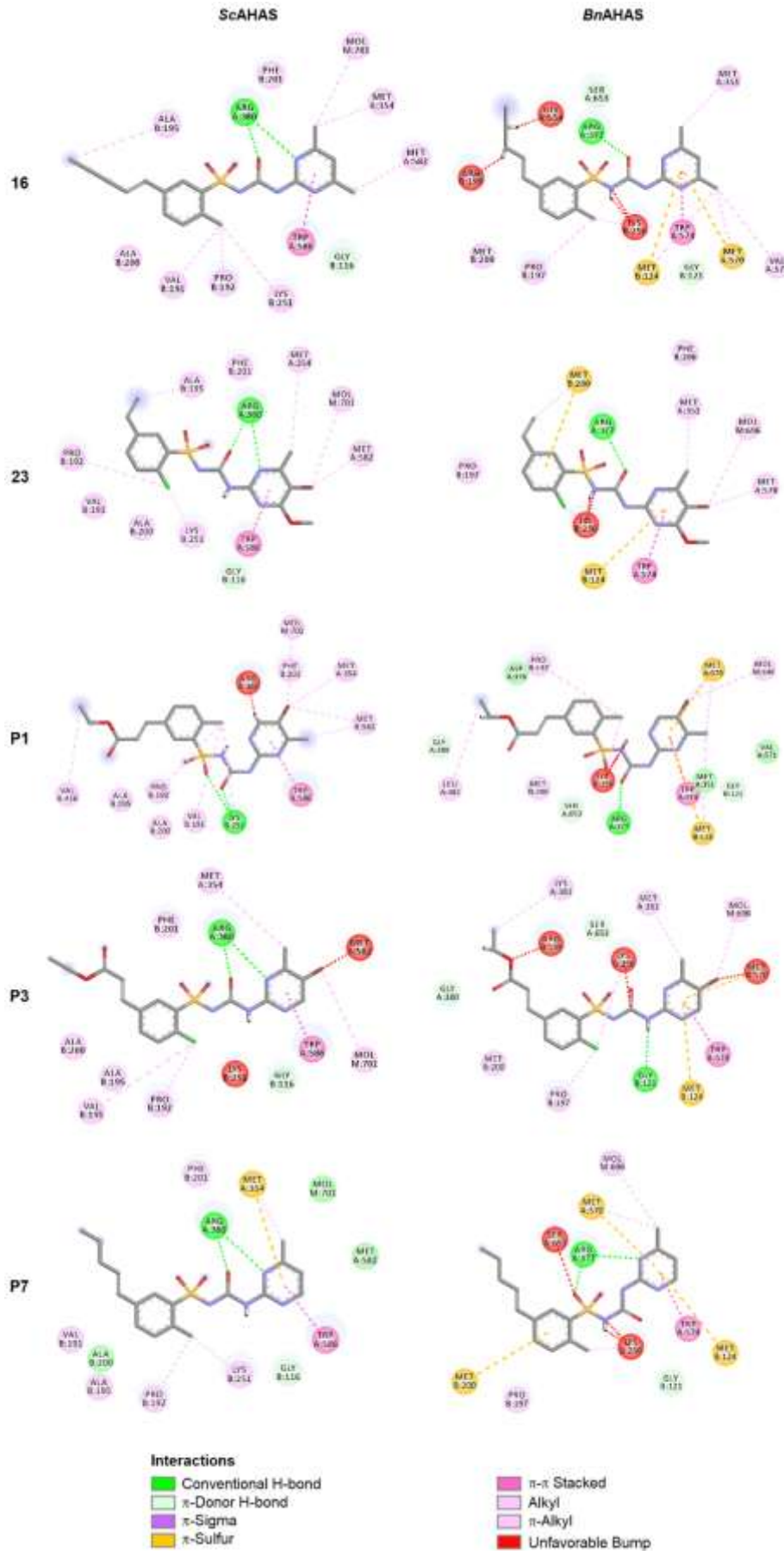


Figure 7. Important ligand-enzyme interactions obtained from the docking studies.

In summary, the MIA-QSAR technique associated with molecular docking studies was employed herein with the purpose of rationally designing novel alkenyl-substituted sulfonylurea derivatives with enhanced antifungal activity and limited herbicidal property, for forage conservation purposes in silage. From the constructed contour maps, it was observed that the modulation of these antagonistic properties could be achieved by particularly modifying the substituents at X (C–Br), R³ (-COOEt, -Et, and -Pr), and R⁵ (Me) positions. As a result of the combination of these substructures, 9 out of 46 novel agrochemical candidates presented very promising calculated selectivity indexes. Docking studies on AHAS corroborated these findings by pointing out important interactions between the proposed ligands and the biological target in question.

ACKNOWLEDGMENTS

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, funding code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant number 301371/2017-2), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

REFERENCES

1. Duggleby, R.G.; McCourt, J.A.; Guddat, L.W. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiol. Biochem.* **2008**, *46*, 309–324.
2. Lee, Y.-T.; Cui, C.-J.; Chow, E.W.L.; Pue, N.; Lonhienne, T.; Wang, J.-G.; Fraser, J.A.; Guddat, L.W. Sulfonylureas have antifungal activity and are potent inhibitors of *Candida albicans* acetohydroxyacid synthase. *J. Med. Chem.* **2013**, *56*, 210-219.
3. Chen, W.; Li, Y.; Shi, Y.; Wei, W.; Chen, Y.; Li, Y.; Liu, J.; Li, B.; Li, Z. Synthesis and evaluation of novel N-(4'-arylpyrimidin-2'-yl) sulfonylurea derivatives as potential antifungal agents. *Chem. Res. Chin. Univ.* **2015**, *31*, 218–223.
4. Chen, W.; Li, Y.; Zhou, Y.; Ma, Y.; Li, Z. Design, synthesis and SAR study of novel sulfonylurea derivatives containing arylpyrimidine moieties as potential anti-phytopathogenic fungal agents. *Chin. Chem. Lett.* **2019**, *30*, 2160–2162.
5. Zhou, S.; Meng, F.-F.; Hua, X.-W.; Li, Y.-H.; Liu, B.; Wang, B.-L.; Chen, J.; Chen, A.-L.; Li, Z.-M. Controllable soil degradation rate of 5-substituted sulfonylurea herbicides as novel AHAS inhibitors. *J. Agric. Food Chem.* **2020**, *68*, 3017–3025.

6. Wei, W.; Cheng, D.; Liu, J.; Li, Y.; Ma, Y.; Li, Y.; Yu, S.; Zhang, X.; Li, Z. Design, synthesis and SAR study of novel sulfonylureas containing an alkenyl moiety. *Org. Biomol. Chem.* **2016**, *14*, 8356-8366.
7. Gramatica, P.; Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: a clarification concerning metrics and terminology. *J. Chem. Inf. Model.* **2016**, *56*, 1127 – 1131.
8. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29*, 476–488.
9. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9*, 91–102.
10. Barigye, S.J.; Freitas, M.P. Ten years of the MIA-QSAR strategy: Historical development and applications. *Int. J. Quant. Struct. Prop. Relat.* **2016**, *1*, 64–77.
11. Geladi, P.; Kowalski, B.R. Partial least squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1-17.
12. Borges, C.N.; Barigye, S.J.; Freitas, M.P. Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients. *Mol. Phys.* **2017**, *115*, 3044-3050.
13. Barigye, S.J.; Duarte, M.H.; Nunes, C.A.; Freitas, M.P. MIA-plot: A graphical tool for viewing descriptor contributions in MIA-QSAR. *RSC Adv.* **2016**, *6*, 49604-49612.
14. Finney, D.J. *Probit analysis*, 2nd ed; Cambridge University Press: New York, 1952.
15. Dennington, R.D.; Keith, T.A.; Millam, M.J. GaussView 5.0, Wallingford, 2008.
16. Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Sys.* **2015**, *145*, 22-29.
17. Mitra, I.; Saha, A.; Roy, K. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol. Simul.* **2010**, *36*, 1067-1079.
18. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on de definition of the Q^2 parameter for QSAR validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669-1678.
19. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N. Some case studies on application of “ r_m^2 ” metrics for judging quality of quantitative structure–activity

- relationship predictions: Emphasis on scaling of response data. *J. Comput. Chem.* **2013**, *34*, 1071-1082.
20. Roy, K.; Das, R.N.; Ambure, P.; Aher, R.B. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Sys.* **2016**, *145*, 22-29.
 21. Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graph. Modell.* **2002**, *20*, 269-276.
 22. Chirico, N.; Gramatica, P. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* **2011**, *51*, 2320-2335.
 23. Chirico, N.; Gramatica, P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.* **2012**, *52*, 2044-2058.
 24. Roy, K.; Mitra, I.; Kar, S.; Ojha, P.K.; Das, R.N.; Kabir, H. Comparative studies on some metrics for external validation of QSPR models. *J. Chem. Inf. Model.* **2012**, *52*, 396-408.
 25. Kiralj, R.; Ferreira, M.M.C. Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *J. Braz. Chem. Soc.* **2009**, *20*, 770-787.
 26. Nunes, C.A.; Freitas, M.P.; Pinheiro, A.C.M.; Bastos, S.C. Chemoface: A novel free user-friendly interface for chemometrics. *J. Braz. Chem. Soc.* **2012**, *23*, 2003-2010.
 27. McCourt, J.A.; Pang, S.S.; Guddat, L.W.; Duggleby, R.G. Elucidating the specificity of binding of sulfonylurea herbicides to acetohydroxyacid synthase. *Biochemistry* **2005**, *44*, 2330-2338.
 28. Fiser, A. Template-based protein structure modeling. In *Computational Biology. Methods in Molecular Biology (Methods and Protocols)*; Fenyö, D., Ed.; Humana Press: Totowa, 2010, Vol. 673, pp. 73-94.
 29. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480-D489.
 30. Huang, X.; Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **1991**, *12*, 337-357.
 31. Hooft, R.W.W.; Vriend, G.; Sander, C.; Abola, E.E. Errors in protein structures. *Nature* **1996**, *381*, 272.

32. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
33. Thomsen, R.; Christensen, M.H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.
34. Ramalho, T.C.; França, T.C.C.; Rennó, M.N.; Guimarães, A.P.; da Cunha, E.F.F.; Kuca, K. Development of new acetylcholinesterase reactivators: molecular modeling versus in vitro data. *Chem. Biol. Interact.* **2010**, *185*, 73-77.
35. R 3.5.3 Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2019. URL (<http://www.R-project.org/>).
36. Orry, A.J.W.; Abagyan, R. *Homology Modeling: Methods and Protocols*. Springer: New York, 2012.
37. Wang, J.-G.; Lee, P.K.-M.; Dong, Y.-H.; Pang, S.S.; Duggleby, R.G.; Li, Z.-M.; Guddat, L.W. Crystal structures of two novel sulfonylurea herbicides in complex with *Arabidopsis thaliana* acetohydroxyacid synthase. *FEBS J.* **2019**, *276*, 1282-1290.
38. Garcia, M.D.; Wang, J.-G.; Lonhienne, T.; Guddat, L.W. Crystal structure of plant acetohydroxyacid synthase, the target for several commercial herbicides. *FEBS J.* **2017**, *284*, 2037-2051.

APPENDIX A – SUPPORTING INFORMATION

**COMPUTER-ASSISTED IMPROVEMENT OF
SULFONYLUREAS WITH ANTIFUNGAL PROPERTIES AND
LIMITED HERBICIDAL ACTIVITY: POTENTIAL
APPLICATION IN FORAGE CONSERVATION**

Adriana Cássia de Faria,^a Joyce K. Daré,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

^a *Departamento de Química, Instituto de Ciências Naturais, Universidade Federal de Lavras, 37200-900, Lavras, MG, Brazil.*

* Corresponding author: matheus@ufla.br

Page S2. **Figure S1.** Applicability domain tests.

Page S3. **Table S1.** Experimental and predicted pEC₅₀ for the MIA-QSAR models without test set.

Page S4. **Table S2.** Experimental and predicted pEC₅₀ for the MIA-QSAR models with test set.

Page S5. **Table S3.** Experimental and predicted *I* for the MIA-QSAR models without test set.

Page S6. **Table S4.** Experimental and predicted *I* for the MIA-QSAR models with test set.

Page S7. **Figure S2.** MIA-plots obtained from MIA descriptors encoding r_{vdW} , ϵ , and r_{vdW}/ϵ , and related to antifungal (pEC₅₀) and herbicidal (*I*) activities.

Page S8. **Figure S3.** Sequence alignment used in the homology modeling. Target sequence: *BnAHAS*. Template sequences: 3EA4. The line with “*” indicates regions of similarity between the two sequences (89.0% identity).

Page S9. **Figure S4.** Docking score energies for the compound data set and proposals.

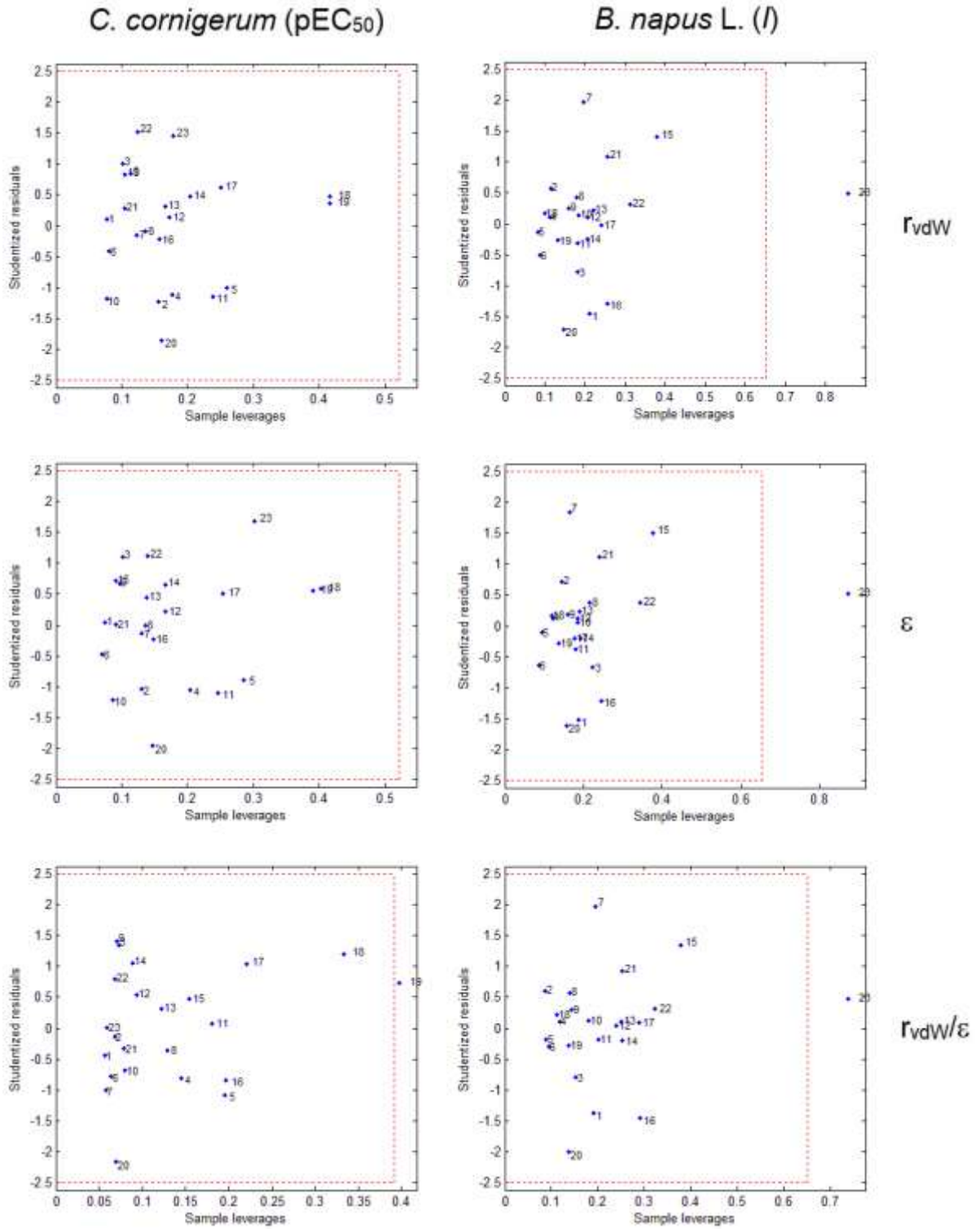


Figure S1. Applicability domain tests.

Table S1. Experimental and predicted pEC₅₀ for the MIA-QSAR models without test set.

Compound	Exp.	Cal.	LOOCV	Cal.	LOOCV	Cal.	LOOCV
		r_{vdw}		ϵ		r_{vdw}/ϵ	
1	1.655	1.661	1.669	1.658	1.659	1.622	1.631
2	1.644	1.570	1.601	1.577	1.570	1.634	1.632
3	1.596	1.658	1.751	1.688	1.744	1.695	1.746
4	1.612	1.546	1.517	1.547	1.478	1.555	1.522
5	1.688	1.632	1.597	1.636	1.568	1.614	1.600
6	1.644	1.618	1.588	1.612	1.614	1.586	1.561
7	1.758	1.749	1.728	1.749	1.731	1.683	1.652
8	1.775	1.770	1.790	1.774	1.796	1.749	1.747
9	1.651	1.703	1.776	1.695	1.749	1.754	1.805
10	1.865	1.791	1.781	1.786	1.762	1.815	1.800
11	1.877	1.812	1.839	1.811	1.796	1.882	1.894
12	1.743	1.751	1.757	1.756	1.758	1.781	1.760
13	1.818	1.837	1.817	1.846	1.845	1.841	1.842
14	1.830	1.858	1.910	1.871	1.933	1.907	1.944
15	1.751	1.802	1.842	1.798	1.840	1.784	1.737
16	1.903	1.899	1.854	1.888	1.844	1.845	1.790
17	1.860	1.895	1.761	1.890	1.781	1.929	1.747
18	1.063	1.086	1.121	1.094	1.152	1.137	1.227
19	1.154	1.172	1.171	1.184	1.201	1.197	1.232
20	1.763	1.653	1.581	1.639	1.554	1.604	1.604
21	1.592	1.609	1.600	1.593	1.615	1.567	1.569
22	1.428	1.521	1.747	1.499	1.736	1.486	1.670
23	1.756	1.842	1.853	1.853	1.863	1.756	1.783

Table S2. Experimental and predicted pEC₅₀ for the MIA-QSAR models with test set.

Compound	Exp.	r_{vdw}			ϵ			r_{vdw}/ϵ		
		Cal.	LOOCV	Pred.	Cal.	LOOCV	Pred.	Cal.	LOOCV	Pred.
1	1.655	1.649	1.673		1.658	1.659		1.627	1.627	
2	1.644	1.568	1.594		1.575	1.553		1.632	1.628	
3	1.596	1.658	1.768		1.661	1.760		1.712	1.777	
4^a	1.612			1.505			1.505			1.534
5	1.688	1.596	1.532		1.591	1.486		1.612	1.564	
6	1.644	1.613	1.589		1.615	1.612		1.586	1.560	
7	1.758	1.740	1.719		1.745	1.728		1.707	1.675	
8	1.775	1.783	1.823		1.790	1.817		1.779	1.769	
9^a	1.651			1.705			1.705			1.758
10^a	1.865			1.796			1.796			1.838
11	1.877	1.838	1.862		1.826	1.792		1.910	1.901	
12	1.743	1.773	1.779		1.775	1.756		1.775	1.745	
13	1.818	1.863	1.856		1.861	1.850		1.853	1.835	
14^a	1.830			1.905			1.905			1.926
15	1.751	1.783	1.892		1.778	1.850		1.785	1.741	
16	1.903	1.874	1.850		1.865	1.828		1.864	1.846	
17	1.860	1.885	1.771		1.879	1.784		1.907	1.747	
18	1.063	1.081	1.126		1.086	1.102		1.110	1.134	
19	1.154	1.171	1.148		1.172	1.171		1.188	1.182	
20	1.763	1.649	1.573		1.646	1.557		1.605	1.564	
21	1.592	1.612	1.599		1.603	1.615		1.564	1.570	
22	1.428	1.519	1.747		1.505	1.739		1.488	1.687	
23	1.756	1.813	1.878		1.836	1.865		1.763	1.773	

^a Test set compounds (the same of Wei et al., 2016).

Table S3. Experimental and predicted *I* for the MIA-QSAR models without test set.

Compound	Exp.	Cal.	LOOCV	Cal.	LOOCV	Cal.	LOOCV
		r_{vdw}		ϵ		r_{vdw}/ϵ	
1	0.811	0.701	0.645	0.695	0.637	0.706	0.653
2	0.776	0.821	0.832	0.832	0.847	0.824	0.837
3	0.638	0.578	0.517	0.588	0.516	0.576	0.519
4	0.453	0.461	0.483	0.463	0.484	0.461	0.485
5	0.230	0.219	0.220	0.221	0.219	0.215	0.217
6	0.917	0.876	0.821	0.865	0.797	0.892	0.850
7	0.308	0.458	0.565	0.451	0.559	0.458	0.560
8	0.329	0.362	0.427	0.357	0.422	0.374	0.432
9	0.711	0.731	0.737	0.725	0.728	0.734	0.740
10	0.477	0.488	0.486	0.481	0.475	0.486	0.486
11	0.416	0.392	0.349	0.387	0.347	0.402	0.356
12	0.418	0.425	0.440	0.427	0.441	0.421	0.434
13	0.167	0.184	0.197	0.185	0.198	0.175	0.187
14	0.106	0.087	0.070	0.091	0.084	0.091	0.069
15	0.279	0.373	0.507	0.379	0.526	0.369	0.489
16	0.225	0.130	0.063	0.136	0.068	0.121	0.048
17	0.514	0.512	0.390	0.499	0.411	0.520	0.385
18	0.525	0.538	0.542	0.538	0.548	0.542	0.542
19	0.318	0.297	0.254	0.296	0.258	0.296	0.255
20	0.975	0.840	0.752	0.848	0.759	0.817	0.722
21	0.936	1.016	1.040	1.019	1.045	1.004	1.021
22	0.925	0.947	0.898	0.951	0.901	0.946	0.935
23	0.075	0.091	0.649	0.091	0.679	0.095	0.625

Table S4. Experimental and predicted *I* for the MIA-QSAR models with test set.

Compound	Exp.	r_{vdW}			ϵ			r_{vdW}/ϵ		
		Cal.	LOOCV	Pred.	Cal.	LOOCV	Pred.	Cal.	LOOCV	Pred.
1	0.811	0.700	0.638		0.695	0.629		0.753	0.736	
2	0.776	0.837	0.870		0.844	0.882		0.824	0.830	
3	0.638	0.586	0.497		0.591	0.493		0.527	0.468	
4^a	0.453			0.486			0.486			0.548
5	0.230	0.236	0.250		0.235	0.245		0.253	0.268	
6	0.917	0.877	0.822		0.867	0.794		0.900	0.855	
7	0.308	0.449	0.562		0.442	0.556		0.456	0.541	
8	0.329	0.361	0.465		0.366	0.472		0.373	0.455	
9^a	0.711			0.745			0.718			0.731
10^a	0.477			0.484			0.465			0.434
11	0.416	0.396	0.317		0.389	0.309		0.351	0.279	
12	0.418	0.414	0.437		0.422	0.454		0.462	0.458	
13	0.167	0.164	0.170		0.171	0.182		0.166	0.180	
14^a	0.106			0.076			0.095			0.083
15	0.279	0.383	0.550		0.386	0.558		0.475	0.583	
16	0.225	0.133	0.045		0.133	0.042		0.178	0.167	
17	0.514	0.516	0.386		0.506	0.407		0.477	0.399	
18	0.525	0.540	0.566		0.539	0.572		0.566	0.594	
19	0.318	0.290	0.248		0.287	0.250		0.270	0.207	
20	0.975	0.843	0.750		0.850	0.756		0.808	0.741	
21	0.936	1.021	1.063		1.022	1.068		0.955	0.910	
22	0.925	0.947	0.905		0.947	0.906		0.957	0.814	
23	0.075	0.089	0.639		0.088	0.667		0.030	0.610	

^a Test set compounds (the same of Wei et al., 2016).

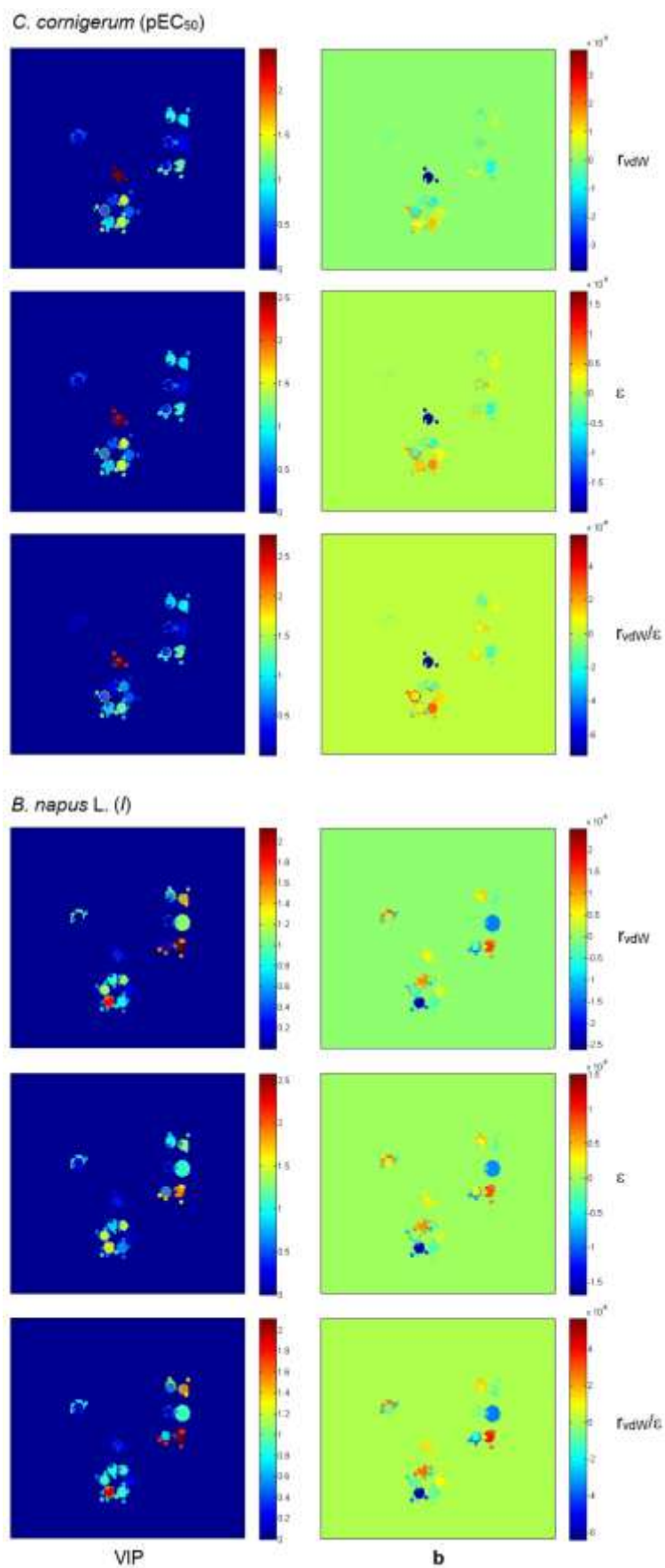


Figure S2. MIA-plots obtained from MIA descriptors encoding r_{vdW} , ϵ , and r_{vdW}/ϵ , and related to antifungal (pEC₅₀) and herbicidal (*I*) activities.

```

3EA4          9  TTSSSISFSTKPSPPSSKSPLPISRFLPFLSLNPNKSSSSRRRGIKSSSPSSISAVLNT
BnAHAS       5  TSSSPISLTAKP---SSKSPLPISRFLPFLSLTPQKDSRLHR-----PLAISAVLNS
              *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
3EA4          69 TTNVTTTPSPTKPTKPKETFISRFPDQPRKGADILVEALERQGVETVFAYPPGASMEIHQ
BnAHAS       55 PVNVAP-PSPEKTDKNKTFVSRYPDEPRKGADILVEALERQGVETVFAYPPGASMEIHQ
              **   ***  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
3EA4          129 ALTRSSSIRNVLPRHEQGGVFAAEGYARSSGKPGICIATSGPGATNLVSGLADALLDSVP
BnAHAS       114 ALTRSSSIRNVLPRHEQGGVFAAEGYARSSGKPGICIATSGPGATNLVSGLADAMLDLSDVP
              *****
3EA4          189 LVAITGQVPRRMIGTDAFQETPIVEVTRSITKHNYLVMDVEDIPRIIEEAFFLATSGRPG
BnAHAS       174 LVAITGQVPRRMIGTDAFQETPIVEVTRSITKHNYLVMDVDDIPRIVQEAFFLATSGRPG
              *****
3EA4          249 PVLVDVPKDIQQQLAIPNWEQAMRLPGYMSRMPKPPEDSHLEQIVRLISESKRPVLYVGG
BnAHAS       234 PVLVDVPKDIQQQLAIPNWDQPMRLPGYMSRLPQPPEVSQLGQIVRLISESKRPVLYVGG
              *****
3EA4          309 GCLNSSDELGRFVELTGIPVASTLMGLGSPYPCDDELSLHMLGMHGTVYANYAVEHSDLLL
BnAHAS       294 GCLNSSDELGRFVELTGIPVASTLMGLGSPYPCDDELSLQMLGMHGTVYANYAVEHSDLLL
              *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
3EA4          369 AFGVRFDDRVTGKLEAFASRAKIVHIDIDSAEIGKNKTPHVSVCGDVKLALQGMNKVLEN
BnAHAS       354 AFGVRFDDRVTGKLEAFASRAKIVHIDIDSAEIGKNKTPHVSVCGDVKLALQGMNKVLEN
              *****
3EA4          429 RAEELKLDFGVWRNELNVQKQKFPKLSFKTFGEAIPPQYAIKVLDELTDGKAIISTGVGQH
BnAHAS       414 RAEELKLDFGVWRSELSEKQKQKFPKLSFKTFGEAIPPQYAIQILDELTEGKAIISTGVGQH
              *****
3EA4          489 QMWAAQFYNYKKPRQWLSSGGLGAMGFLPAAIGASVANPDIAIVVDIDGDGFSIMNVQEL
BnAHAS       474 QMWAAQFYKYRKRQWLSSSGLGAMGFLPAAIGASVANPDIAIVVDIDGDGFSIMNVQEL
              *****
3EA4          549 ATIRVENLPVKVLLLNQHLGMVMQWEDRFYKANRAHTFLGDPAQEDEFIPNMLLFAAAC
BnAHAS       534 ATIRVENLPVKIILLNQHLMGMVMQWEDRFYKANRAHTYLGDPARENEIFPNMLQFAGAC
              *****
3EA4          609 GIPAARVTKKADLREAIQTMLDTPGPYLLDVICPHQEHVLPMPISGGTFNDVITEGDGRI
BnAHAS       594 GIPAARVTKKEELREAIQTMLDTPGPYLLDVICPHQEHVLPMPISGGTFKDVITEGDGRT
              *****
3EA4          669 KY
BnAHAS       654 KY
              **

```

Figure S3. Sequence alignment used in the homology modeling. Target sequence: *BnAHAS*. Template sequences: 3EA4. The line with “*” indicates regions of similarity between the two sequences (89.0% identity).



Figure S4. Docking score energies for the compound data set and proposals.