



NATÂNIA ESTEFANY RODRIGUES

**STUDY OF THE HERBICIDAL PROFILE AND SOIL
SORPTION OF SULFONYLUREAS BY MOLECULAR
MODELING**

LAVRAS-MG

2023

NATÂNIA ESTEFANY RODRIGUES

**STUDY OF THE HERBICIDAL PROFILE AND SOIL SORPTION OF
SULFONYLUREAS BY MOLECULAR MODELING**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração Química/Bioquímica, para o título de Mestre.

Prof. Dr. Matheus Puggina de Freitas
Orientador

LAVRAS-MG

2023

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

Rodrigues, Natânia Estefany.

Study of the herbicidal profile and soil sorption of sulfonylureas
by molecular modeling / Natânia Estefany Rodrigues. - 2023.

90 p.

Orientador(a): Matheus Puggina de Freitas.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2023.

Bibliografia.

1. Sulfonylurea. 2. MIA-QDAR. 3. Herbicidas. I. de Freitas,
Matheus Puggina. II. Título.

NATÂNIA ESTEFANY RODRIGUES

**STUDY OF THE HERBICIDAL PROFILE AND SOIL SORPTION OF
SULFONYLUREAS BY MOLECULAR MODELING**

Dissertação apresentada à
Universidade Federal de Lavras,
como parte das exigências do
Programa de Pós-Graduação em
Agroquímica, área de concentração
Química/Bioquímica, para o título
de Mestre.

APROVADA em 19 de janeiro de 2023.

Dr. Sérgio Scherrer Thomasi UFLA

Dr. João Eustáquio Antunes UFJF

Prof. Dr. Matheus Puggina de Freitas

Orientador

LAVRAS-MG

2023

AGRADECIMENTOS

Agradeço primeiramente à Deus que me presenteou com o bem mais precioso que eu poderia receber, a vida, e com ela a capacidade lutar pelas conquistas de meus ideais;

Aos meus pais, Antônio Carlos, Rosa Inês e meu padrasto Vitor Ozair, minhas maiores inspirações, por sempre acreditarem no meu potencial e não medirem esforços para me auxiliarem ao longo dessa jornada. Especialmente à minha mãe, que sempre me mostrou os caminhos certos a seguir reforçando a importância de estudar, e por me auxiliar a ser quem eu sou, serei eternamente grata;

Às minhas irmãs, Jéssica e Natacha, que são meus tesouros pela força e apoio em todos os momentos;

Ao meu marido, Paulo Cezar por todo apoio, por me erguer sempre que achava que não era capaz, você foi essencial;

Ao meu orientador, Matheus Puggina de Freitas pela excelente orientação, apoio, ensinamentos, paciência e incentivo para a realização deste trabalho, serei eternamente grata;

Agradeço aos professores da banca examinadora pela disponibilidade e contribuições para este trabalho;

Aos meus amigos que sempre torceram para o meu sucesso, às amigas que foram construídas ao longo do curso, com quem sempre compartilhei alegrias, tristezas, frustrações, satisfações e ensinamentos;

Ao programa de pós-graduação em Agroquímica por todo apoio;

À todos colegas e amigos do laboratório de Química Computacional;

À todos os professores do DQI/UFLA por compartilhar seus ensinamentos;

À Universidade Federal de Lavras, pela oportunidade de realizar o curso de mestrado e por todo subsídio;

O presente trabalho foi realizado com o apoio à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pela concessão da bolsa de mestrado.

RESUMO

Plantas daninhas podem provocar grandes prejuízos à agricultura. Portanto, o desenvolvimento de novos herbicidas é necessário para garantir a conservação de culturas agrícolas e, ao mesmo tempo, servir de alternativa para o controle de pragas resistentes. Derivados de sulfonilureias são compostos amplamente empregados como herbicidas na agricultura. Uma técnica computacional utilizada no *design* de novos agroquímicos, bem como na interpretação dos aspectos físico-químicos que explicam uma determinada resposta biológica, é o QSAR (*Quantitative Structure-Activity Relationships*). Dentre os métodos existentes, o presente trabalho utilizou a técnica MIA-QSAR, que consiste na análise multivariada de imagens aplicada em QSAR, para investigar como a variação nos substituintes de uma série congênere de moléculas previamente sintetizadas e testadas contra a *Brassica napus L.* (derivados de pirimidil acil tiureias e derivados de sulfonilureias contendo grupos aril 2,6-dissubstituídos) explica a variação nos dados das atividades herbicidas. O modelo para os derivados de pirimidil acil tiureias mostrou-se preditivo por apresentar parâmetros estatísticos de validação significativos, como o r^2_{pred} de 0,833 e r^2 de 0,829. Três novos herbicidas foram propostos *via* método *in silico* com características promissoras. Já para os herbicidas contendo grupos aril 2,6-dissubstituídos, os modelos desenvolvidos apresentaram um bom desempenho $r^2 > 0,90$, bem como alto coeficiente de determinação na validação cruzada *leave-one-out* $q^2 > 0,76$ e, além disso, quatro compostos foram propostos, sendo que três desses apresentaram atividade herbicida maior que os da literatura. Estudos de *docking* molecular também foram realizados para compreender as interações ligante-enzima responsáveis pela atividade dos herbicidas. Um estudo SAR da sorção do solo de herbicidas (sulfonilureias) conhecidos e comercializados foi desenvolvido, utilizando valores de $\log K_{OC}$ experimentais. Descritores de análise de imagem multivariada (MIA) foram empregados para construir modelos da propriedade de estrutura para classificar esses compostos de acordo com sua capacidade de sorção do solo. Além disso, uma análise mais profunda baseada em mapas de contorno MIA de coeficientes de regressão PLS e importância variável em pontuações de projeção foi realizada para obter informações sobre as características químicas responsáveis pelo comportamento de $\log K_{OC}$. Um modelo de regressão linear múltipla obtido de descritores selecionados demonstrou alta previsibilidade ($r^2 = 0,95$, $q^2 = 0,84$ e $r^2_{\text{pred}} = 0,71$) e também forneceu informações químicas para os valores de $\log K_{OC}$. Verificou-se que os derivados de triazina, em vez de pirimidina, são menos propensos a lixiviação, mas os substituintes ligados ao grupo sulfonil desempenham um papel importante na modulação dos valores de $\log K_{OC}$.

Palavras-chaves: Sulfonilureia. MIA-QSAR. Herbicidas. $\log K_{OC}$.

ABSTRACT

Weeds can cause great damage to agriculture. Therefore, the development of new herbicides is necessary to guarantee the conservation of agricultural crops and, at the same time, serve as an alternative for the control of resistant pests. Derivatives of sulfonyleureas are compounds widely used as herbicides in agriculture. A computational technique used in the design of new agrochemicals, as well as in the interpretation of physical-chemical aspects that explain a certain biological response, is the QSAR (*Quantitative Structure-Activity Relationships*). Among the existing methods, the present work used the MIA-QSAR technique, which consists of the multivariate analysis of images applied in QSAR, to investigate how the variation in the substituents of a congeneric series of molecules previously synthesized and tested against *Brassica napus* L. (pyrimidyl acyl thiourea derivatives and sulfonyleurea derivatives containing 2,6-disubstituted aryl groups) explains the variation in the herbicidal activity data. The model for pyrimidyl acyl thioureas derivatives proved to be predictive by presenting significant validation statistics parameters, such as r^2_{pred} of 0.833 and r^2 of 0.829. Three new herbicides were proposed via in silico method with promising characteristics. As for the herbicides containing 2,6-disubstituted aryl groups, the developed models showed a good performance $r^2 > 0.90$, as well as a high coefficient of determination in the cross-validation leave-one-out $q^2 > 0.76$ and, in addition, four compounds were proposed, and three of these showed herbicidal activity greater than those in the literature. Molecular *docking* studies were also performed to understand the ligand-enzyme interactions responsible for herbicide activity. A SAR study of soil sorption of known and commercially available herbicides (sulfonyleureas) was carried out using experimental $\log K_{OC}$ values. Multivariate image analysis (MIA) descriptors were employed to build structure property models to classify these compounds according to their soil sorption capacity. Furthermore, a deeper analysis based on MIA contour maps of PLS regression coefficients and variable importance in projection scores was performed to obtain information about the chemical characteristics responsible for $\log K_{OC}$ behavior. A multiple linear regression model obtained from selected descriptors demonstrated high predictability ($r^2 = 0.95$, $q^2 = 0.84$ and $r^2_{\text{pred}} = 0.71$) and also provided chemical information for $\log K_{OC}$ values. It was found that triazine derivatives, rather than pyrimidine, are less prone to leaching, but substituents attached to the sulfonyl group play an important role in modulating $\log K_{OC}$ values.

Keywords: Sulfonyleurea. MIA-QSAR. Herbicides. $\log K_{OC}$.

SUMÁRIO

	PRIMEIRA PARTE	9
1	INTRODUÇÃO	10
2	JUSTIFICATIVA	13
3	OBJETIVOS	14
3.1	Objetivo Geral	14
3.2	Objetivos específicos	14
4	REFERENCIAL TEÓRICO	15
4.1	Herbicidas	15
4.2	AHAS	16
4.3	Sulfonilureias	17
4.4	QSAR (Quantitative Structure-Activity Relationships)	19
4.5	Multivariate Image Analysis applied to QSAR (MIA-QSAR)	21
4.6	Análise multivariada	25
4.6.1	MLR - Regressão Linear Múltipla	26
4.6.2	PLS - Mínimos Quadrados Parciais	26
4.6.3	Validação QSAR	27
5	DOCKING MOLECULAR	29
6	MATERIAIS E MÉTODOS	31
	REFERÊNCIAS BIBLIOGRÁFICAS	40
	SEGUNDA PARTE	47
	ARTIGO 1	48
	ARTIGO 2	63
	ARTIGO 3	77

PRIMEIRA PARTE

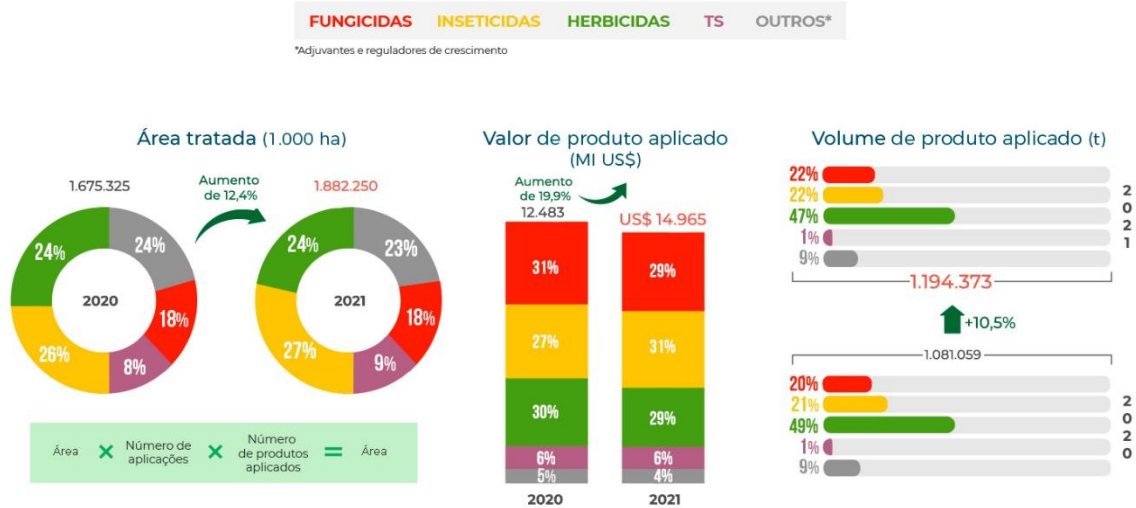
1 INTRODUÇÃO

A exportação de *commodities*, sobretudo de produtos ligados à agricultura, é a principal atividade econômica do Brasil. As plantações de soja, milho e algodão correspondem por cerca de 90% da área plantada com grãos e cereais, 35% das exportações do agronegócio e 16% dos empregos na agricultura brasileira (VON ZUBEN, 2019). Porém, devido ao rápido crescimento e facilidade de disseminação, as plantas daninhas podem ser responsáveis por grandes perdas de produtividade. As plantas daninhas são reconhecidas como problemas para a produção agrícola, conferindo limitações significativas para alcançar potenciais rendimentos em, por exemplo, lavouras de cana-de-açúcar e arroz (CHIAPINOTTO *et al.*, 2017; RODRÍGUEZ; BARBOSA; RODRÍGUEZ, 2019). Com isso, as plantas daninhas podem competir diretamente com as safras por recursos físicos, como luz, e por recursos hídricos, além de comprometer a qualidade da safra por meio da contaminação, atuando como um reservatório biológico de pragas e doenças. As plantas daninhas podem ocasionar perdas expressivas na produção, em média de 36%, e em alguns casos pode-se chegar a mais de 80% de perda (KARAM *et al.*, 2018). Por exemplo, no caso da produção de grãos, as perdas ficam, em média, entre 13 e 15% das safras e podem chegar à casa dos 90% em lavouras de outros cultivares (HENRIQUE, 2019).

A fim de maximizar a produção, são adotadas e implementadas uma gama de estratégias de controle e, dentre elas, pode-se apontar a mecânica, a biológica e a química. As estratégias adotadas variam de acordo com as particularidades da safra e a mais comumente utilizada por agricultores para o controle das ervas daninhas é a química, por meio de herbicidas. Esse fato se dá em razão da eficiência, praticidade e economia em comparação a outros métodos de controle (KARAM *et al.*, 2018).

Os pesticidas mais utilizados nas plantações do Brasil são os herbicidas (Figura 1). Dentre as classes de herbicidas, destacam-se os inibidores de acetolactato sintase (ALS, ou aceto-hidroxiácido sintase – AHAS), os inibidores do fotossistema II (FSII), os inibidores de acetil-CoA carboxilase (ACCase), as auxinas sintéticas, e os inibidores da 5-enolpiruvilchiquimato-3-fosfato (EPSP) sintase (KRAEHMER *et al.*, 2014), bem como inibidores da 4-hidroxifenilpiruvato dioxigenase (HPPD).

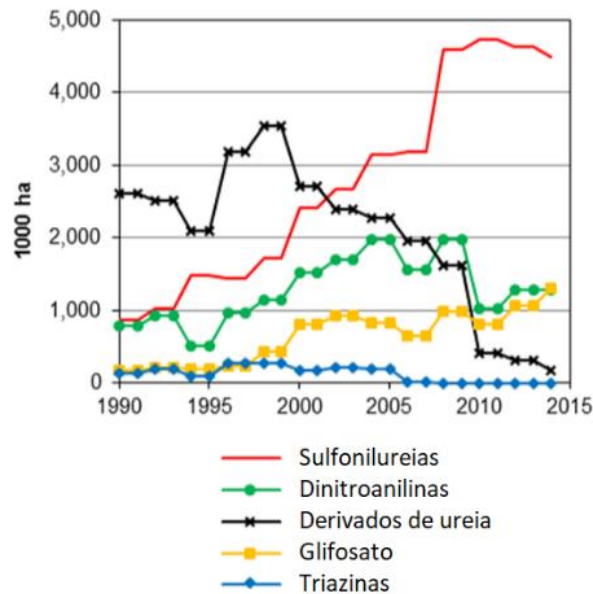
Figura 1- Percentual de defensivos agrícolas aplicado por área tratada, valor e volume do produto aplicado em 2020 e 2021 no Brasil.



Fonte: Sindeveg/ Spark Consultoria (2022).

Ainda dentre os herbicidas, as principais classes de compostos são as triazinas, o glifosato (um organofosforado), os derivados de ureia, as dinitroanilinas e, sobretudo, as sulfonilureias (Figura 2) (ERVIN, 2016).

Figura 2- Classes de herbicidas mais vendidos no mundo.

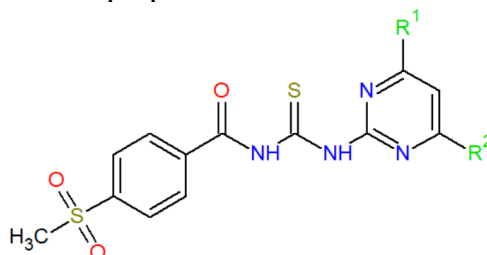


Fonte: ERVIN (2016).

Recentemente, uma série de derivados de sulfonilureias representada pela estrutura da Figura 3, foi sintetizada e alguns representantes dessas classes de compostos foram reportados

como potenciais herbicidas, possivelmente inibidores da AHAS, em várias concentrações (LI *et al.*, 2021a; 2021b; WU *et al.*, 2021). Como se observa (Figura 3), a mudança nos substituintes das moléculas explica os diferentes comportamentos biológicos no conjunto de compostos, uma vez que a unidade básica (a porção acil tiureia) permanece constante.

Figura 3- Estrutura básica de uma sulfonilureias derivada de pirimidil acil tiureias. A variação nos substituintes R altera a propriedade herbicida dos seus derivados.



Fonte: Do autor (2022).

A ferramenta computacional que busca encontrar a relação entre a variação dos substituintes em uma molécula e a variação nas respostas biológicas denomina-se QSAR, do inglês *Quantitative Structure-Activity Relationships*, por meio de descritores moleculares. Os descritores moleculares são os parâmetros, calculados ou experimentais, que estabelecem a relação entre estrutura e atividade por meio de um modelo matemático de regressão. Os descritores moleculares correspondem a quaisquer parâmetros físicos ou químicos que codifiquem aspectos estruturais de uma molécula e que se relacionem com a propriedade de interesse (TODESCHINI; CONSONNI; MANNHOLD, 2000). Na técnica MIA-QSAR (*Multivariate Image Analysis applied to QSAR*), os descritores correspondem aos píxeis de imagens sobreponíveis de moléculas (BARIGYE; DE FREITAS, 2016).

O presente trabalho tem por finalidade encontrar uma correlação entre descritores moleculares MIA de dois conjuntos de derivados de sulfonilureias suas correspondentes atividades herbicidas. O modelo MIA-QSAR a ser construído poderá ser utilizado na previsão da atividade herbicida de potenciais candidatos agroquímicos derivados dessas classes. Para interpretar os resultados químicos, serão construídos e analisados os MIA-Plots. Além disso realizar um estudo SAR da sorção do solo de herbicidas de sulfonilureias conhecidos comercialmente. Docking molecular, também chamado de ancoramento molecular, será empregado para compreender as interações ligante-enzima que explicam as atividades herbicidas observadas.

2 JUSTIFICATIVA

Os herbicidas de sulfonilureias são conhecidos por sua eficácia no controle de ervas daninhas e por apresentar baixa toxicidade. Entretanto, é comum na agricultura que ocorra a resistência das plantas daninhas aos herbicidas amplamente utilizados, consequentemente provocando danos às culturas.

O desenvolvimento de modelos MIA-QSAR é uma ferramenta computacional preditiva e robusta na geração de modelos de previsão de atividades biológicas. Podem ser utilizados para uma série de potenciais herbicidas derivados de sulfonilureias, Tais modelos são ferramentas que possibilitam ampliar o escopo de agroquímicos utilizados no controle de plantas daninhas e que ao mesmo tempo, proponha alternativas para o controle de pragas resistentes aos herbicidas existentes.

3 OBJETIVOS

3.1 Objetivo Geral

Construir modelos MIA-QSAR das atividades herbicidas derivados de sulfonilureias, bem como compreender os aspectos químicos responsáveis pelas suas atividades biológicas por meio de MIA-plots e *docking* molecular.

3.2 Objetivos específicos

- a) Realizar a construção de modelos MIA-QSAR com elevada capacidade preditiva das atividades herbicidas de derivados de sulfonilureias.
- b) Realizar regressão por mínimos quadrados parciais (PLS) dos descritores MIA, com a finalidade de compreender as regiões e substituintes que mais afetam a bioatividade nas moléculas dos herbicidas.
- c) Validar o modelo MIA-QSAR construído e utilizá-lo para a previsão da atividade herbicida de novos candidatos a agroquímicos;
- d) Realizar estudos de *docking* molecular para validar as previsões obtidas *via* MIA-QSAR e compreender o modo de interação entre as moléculas propostas e o seu sítio de ação na enzima AHAS.
- e) Realizar um estudo SAR e QSAR da sorção no solo de herbicidas de sulfonilureias conhecidos através de descritores MIA.
- f) Divulgar os resultados em publicações em revistas com fator de impacto relevante par área.

4 REFERENCIAL TEÓRICO

4.1 Herbicidas

Herbicidas são compostos químicos da classe dos agroquímicos utilizados para controle de plantas daninhas. Esses produtos atuam na seleção de certas populações de plantas, isto é, provocam a morte de determinadas plantas e de outras não. Esses compostos, atualmente, são os mais usados em escala mundial no manejo de plantas indesejáveis (XAVIER *et al.*, 2018).

Esse fato está atrelado ao alto índice de consumo e demanda de alimentos, o que, conseqüentemente, exige dos produtores rurais uma maior produção. Porém, como as plantações são sistemas complexos e orgânicos, estão suscetíveis a inúmeros fatores que podem afetar a qualidade e a disponibilidade das culturas. Por exemplo, as ervas daninhas podem competir com a cultura pela disponibilidade de nutrientes minerais essenciais, por água e luz, bem como por espaço, dentre outros tipos de competição. Assim, faz-se com que o produtor busque por alternativas e meios para controlar as pragas que afetam os cultivares (JARDIM; ANDRADE; QUEIROZ, 2009).

Desse modo, existem alternativas de controle das ervas daninhas, como o controle mecânico, o biológico e químico. O controle químico possui inúmeras vantagens de aplicação, como impedir a interferências das plantas daninhas no início do ciclo, período que acarreta muitas perdas. Além disso, o uso de herbicida proporciona um controle mais efetivo nos plantios, bem como possui uma grande flexibilidade de aplicação, podendo ser aplicado em etapas, a depender da demanda da cultura.

Analogamente aos benefícios, existem também limitações ao uso de herbicidas. Esses apresentam riscos para o homem, para outras plantas, meio ambiente e animais, devido ao grau de toxicidade. O risco a saúde humana está atrelado tanto ao trabalhador rural que manipula o herbicida, quanto aos consumidores que consomem os produtos com resíduos de pesticidas químicos. Diante disso, existe a necessidade da produção de herbicidas ambientalmente amigáveis, com impactos reduzidos à saúde e ao ambiente.

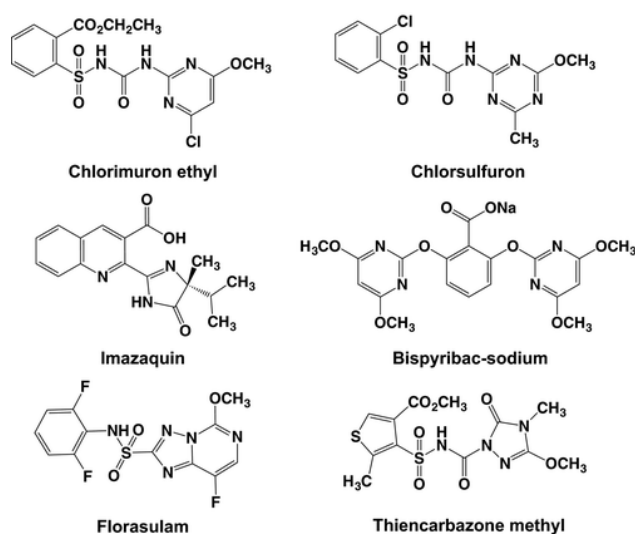
Nos últimos anos têm-se produzidos herbicidas cada vez mais ambientalmente seguros. Desse modo, novos compostos candidatos a herbicidas mais potentes e que apresentem riscos ambientais reduzidos têm sido explorados e, dentre eles os compostos à base de sulfonilureias mostram-se como objetos atraentes, devido à sua elevada atividade inibitória contra a enzima aceto-hidroxiácido sintase (AHAS).

4.2 AHAS

A aceto-hidroxiacido sintase (AHAS), conhecida também como acetolactato sintase (ALS), é a primeira enzima na via da biossíntese de aminoácidos de cadeia ramificada (BCAAs), ou seja, valina, leucina e isoleucina (LIANG *et al.*, 2021; CHIAPINOTTO *et al.*, 2017). A AHAS tem por finalidade realizar a conversão de duas moléculas de piruvato em (*S*)-2-acetolactato, ou converter uma molécula de piruvato e uma molécula de 2-cetobutirato (*S*)-cetobutirato em (*S*)-2-aceto-2-hidroxitubirato (GARCIA, et al., 2017). A enzima AHAS é alvo de algumas classes de herbicidas comerciais. As quais se classificam em 5 diferentes grupos: sulfonilureias (Sus), imidazolinonas (IMIs), pirimidinil (tio)-benzoatos (PTBs), triazolopirimidinas (TPs) e sulfonilamino-carbonil-triazolinona (SCT), conforme ilustrado na Figura 4 (QU *et al.*, 2017; GARCIA *et al.*, 2017)

As classes de herbicidas apresentadas acima são essenciais para o controle das plantas daninhas pela expressiva atividade herbicida, além de serem popularmente conhecidas pela alta seletividade e por apresentar baixa toxicidade em mamíferos, bem como pela baixa concentração de aplicação (MARIANI *et al.*, 2019).

Figura 4- Exemplos das estruturas químicas das cinco classes diferentes de herbicidas AHAS: sulfonilureias (clorimuron-etil e CS), imidazolinonas (imazaquin), pirimidinil-benzoatos (bispiribac-sódio), triazolopirimidinas (florasulam) e sulfonilamino-carbonil-triazolinonas (tiencarbonil-triazolinonas metil).



Fonte: GARCIA *et al.* (2017).

Algumas plantas são suscetíveis ao ataque de herbicidas inibidores da enzima AHAS, que se ligam ao sítio catalítico da enzima, impedindo que o substrato da síntese se conecte a ele, provocando a inibição da síntese de BCAAs e causando a morte da planta (LI, et al., 2018). Desse modo, a AHAS é considerada um objeto de estudo atraente para novos fungicidas, herbicidas e bactericidas (WU *et al.*, 2019).

Entretanto, o uso excessivo a longo prazo desse herbicida e de outros inibidores de AHAS desencadeou um desenvolvimento de resistência das plantas daninhas em aplicações agrícolas no Brasil e no mundo (CECHIN *et al.*, 2016 BRUSAMARELLO *et al.*, 2016; DURIGON; *et al.*, 2018; MENDES *et al.*, 2019). A evolução da resistência está comumente relacionada a uma redução na sensibilidade do sítio ativo, ocasionada por mutações genéticas (PANDOLFO, 2013).

Jiménez-Bolaños e colaboradores (2018) realizaram uma pesquisa para verificar a atividade biológica dos herbicidas da acetil coenzima carboxilase e da acetolactato sintase, bem como a resistência em um tipo específico de planta daninha. Os autores observaram um controle biológico de 77% pelos inibidores da AHAS e foram encontrados indícios de resistência efetiva nos herbicidas inibidores da enzima acetolactato sintase.

No Brasil, o primeiro caso de resistência múltipla foi detectado no estado do Paraná pelo biótipo *Euphorbia heterophylla*, com resistência aos herbicidas inibidores das enzimas protoporfirinogênio oxidase (PROTOX) e acetolactato sintase (ALS). Existem também relatos da ocorrência de biótipos resistentes a enzima ALS da espécie, espécie *Amaranthus palmeri* nas áreas de produção de algodão, no estado de Mato Grosso (GONÇALVES NETTO *et al.*, 2019). Outro exemplo de casos de resistência de plantas daninhas é relatado no Rio Grande do Sul (ULGUIM *et al.*, 2019).

Dentre os herbicidas inibidores de AHAS as sulfonilureias (Sus) começaram a ser amplamente utilizadas desde o ano de 1980, devido à sua alta seletividade, que contribui efetivamente para a variedade de culturas, além de possuir efeitos ambientais benignos, dentre outros fatores que exemplificam a sua utilização. Tornaram-se os herbicidas mais amplamente utilizados no mundo (GUO *et al.*, 2020; ZHOU *et al.*, 2020; LI *et al.*, 2018; CECHIN *et al.*, 2017; WEI *et al.*, 2017).

4.3 Sulfonilureias

Herbicidas de sulfonilureias (Sus) são um grupo amplamente utilizado para controle de uma diversidade de ervas daninhas e gramíneas em diversas culturas, como arroz, milho,

beterraba, cevada e tomate. Devido à alta eficácia em baixas concentrações de aplicação (10-40 g/ha) e baixa toxicidade, o grupo de herbicidas obteve uma acelerada aceitação no mundo todo. A estrutura geral das sulfonilureias é $R_1-SO_2-NH-CO-NH-R_2$, no qual o R_1 pode ser inserido um grupo alifático, aromático ou heterocíclico, enquanto em R_2 insere-se uma triazina ou um anel de pirimidina substituída. (WEI *et al.*, 2017; ZHOU *et al.*, 2020).

Em geral, as sulfonilureias podem ser ativas tanto por via foliar, quanto por via solo. A via foliar consiste em aplicar o herbicida na parte aérea da planta (as folhas), enquanto pela via solo a aplicação se dá diretamente no solo onde se encontra a planta daninha; os herbicidas podem se transcolar pelos caminhos do apoplasto e simplasto. O apoplasto consiste num sistema de absorção que abrange os espaços externos às membranas plasmáticas celulares. Nesse conjunto, o herbicida não se desloca por nenhuma membrana, se dá de forma contínua e não há gasto de energia, já que ocorre difusão simples. O simplasto, em contrapartida, está associado ao conteúdo intracelular. Nesse processo o movimento ocorre no interior das células, através dos tubos citoplasmáticos responsáveis pelo contato direto entre os citoplasmas de células (DE OLIVEIRA; CONSTANTIN; INOUE, 2011).

No ambiente, os herbicidas de sulfonilureias podem ser degradados por hidrólise química e por degradação microbiana (DE OLIVEIRA; CONSTANTIN; INOUE, 2011). São ácidos fracos por possuírem pK_a variando de 3 a 5; em solução aquosa existem na forma neutra quando $pH < pK_a$ e na forma aniônica quando $pH > pK_a$. É de suma importância para fins de proteção e para os sistemas agrônômicos compreender como os herbicidas de sulfonilureias se comportam em relação à sua persistência e mobilidade no solo (SARMAH; SABADIE, 2002).

Alguns herbicidas de Sus são conhecidos por sua persistência no solo, e atividade residual longa, ocasionando prejuízos e provocando danos às plantas frágeis a esses compostos. Esse grupo de compostos é conhecido por ter a degradação no solo dependente do pH (BOSCHIN; *et al.*, 2007). A hidrólise química é facilitada em solos ácidos, enquanto em solos neutros e alcalinos é considerada lenta, conseqüentemente à persistência torna-se maior. Portanto, torna-se importante avaliar, prevenir e controlar a poluição causada por esses compostos químicos.

A sorção de produtos químicos orgânicos em solos desempenha um papel importante de seu transporte e destino ambiental. O termo sorção é empregado para compreender a retenção de forma geral, sem especificar ou distinguir processos específicos como absorção e adsorção (LAVORENTI; PRATA; REGINATO, 2003). Portanto, representa a “captação” de determinado soluto pela matriz do solo. A sorção de herbicidas pelos componentes do solo

influencia na quantidade presente na solução do solo, e, portanto, determina a mobilidade, a lixiviação, persistência e biodisponibilidade no meio.

A potencial de sorção é representado pelo coeficiente de sorção normalizado pela fração carbono orgânico do solo K_{OC} :

$$K_{OC} = \frac{K_D \times 100}{f_{OC}} \quad (1)$$

Onde, K_D é o coeficiente de sorção linear, representando a fração adsorvida do herbicida no solo pela fração de herbicida na solução do solo. Já o termo f_{OC} representa a porcentagem de teor de carbono orgânico do solo. Quanto maior for o K_{OC} , maior será a correlação entre o herbicida e a fração da matéria orgânica do meio. As principais interações envolvidas na sorção de herbicidas pela matéria orgânica são as interações hidrofóbicas, isto é, interações de moléculas orgânicas pela parte orgânica do solo, devido à sua baixa afinidade hidrofílica. Portanto, o coeficiente K_{OC} pode ser um aliado na descoberta de novos compostos bioativos de sulfonilureias ao permitir compreender as características estruturais que afetam as propriedades físicas e ambientais.

Existe uma vasta demanda para o desenvolvimento de novos herbicidas ainda mais potentes (HENCKES *et al.*, 2019; QU *et al.*, 2017). Para isso, faz-se necessário conhecer a estrutura da enzima alvo (AHAS) (MAO *et al.*, 2018; JANA; DELGADO; MEDINA, 2014), além de compreender as interações entre o inibidor e a enzima, para que se avance no desenvolvimento de novos herbicidas. Um aliado, e uma estratégia muito utilizada para o *design* molecular, é o método quantitativo de relação estrutura-atividade (QSAR). O QSAR é uma ferramenta computacional de modelagem molecular que auxilia a descoberta de novos compostos bioativos (LI *et al.*, 2021).

Diversos autores vêm se empenhando no estudo do desenvolvimento de possíveis efeitos a herbicidas utilizando a metodologia QSAR (HUANG *et al.*, 2021; WANG *et al.*, 2020; CHEN *et al.*, 2019; TONG *et al.*, 2019). Com isso, infere-se que QSAR é uma importante ferramenta de pesquisa quando aplicada ao desenvolvimeto de produtos químicos com atividade biológica, como os herbicidas (CARVALHO; FERREIRA, 2019).

4.4 QSAR (Quantitative Structure-Activity Relationships)

QSAR é uma ferramenta computacional de destaque na comunidade científica de modelagem molecular (MARTINS; FERREIRA, 2013). É uma técnica muito utilizada pela área de Química Medicinal para a descoberta de novos fármacos, auxiliando no planejamento e

desenvolvimento de fármacos, visto que permite propor mudanças estruturais que possibilitem a obtenção de moléculas com propriedades farmacológicas (MARTINS; FERREIRA, 2013).

Utiliza-se a estratégia de Planejamento Baseado na Estrutura do Ligante (LBDD, do inglês, *Ligand-Based Drug Design*), que consiste na análise das implicações que as alterações na estrutura do composto (ligante) pode ocasionar na sua respectiva atividade biológica e, por efeito, pode-se explicar os principais motivos que governam as transformações (ARROIO; HONÓRIO; SILVA;2010). Além de ser um método muito importante para a área de Química Medicinal, tornou-se também relevante para diversos campos científicos, como no âmbito da agroquímica, na busca de novos compostos bioativos, tal como os herbicidas (GUIDO; ANDRICOPULO; OLIVA, 2010).

O método QSAR busca identificar e quantificar as relações predominantes entre a estrutura química e atividade biológica de um conjunto de moléculas quimicamente relacionadas (congêneres) através do uso de descritores. Os descritores são índices numéricos que codificam as informações relacionadas à estrutura molecular dos compostos, pois estão intimamente ligados ela e desempenham um papel central para o desenvolvimento do modelo QSAR (TODESCHINI; CONSONNI; MANNHOLD, 2000).

Os modelos QSAR inferem que a estrutura molecular de um composto possui propriedades, informações e características responsáveis pelo seu comportamento físico, químico e biológico, bem como sobre a capacidade de capturar essas características por descritores numéricos. A habilidade de registrar as informações necessárias está intimamente ligada ao tipo de descritor escolhido, dependendo da representação molecular e dos cálculos utilizados (MARTINS; FERREIRA, 2013).

Os descritores moleculares são as variáveis independentes calculadas para cifrarem as informações necessárias a respeito da atividade biológica (variável dependente) do conjunto de moléculas para a construção do modelo. De modo geral, não existe um descritor melhor e absoluto para a tarefa da modelagem molecular na construção de um bom modelo QSAR, e sim uma seleção adequada de descritores moleculares, de acordo com as particularidades de cada conjunto de dados e objetivos do projeto. De acordo com a representação escolhida para a geração dos descritores, eles podem ter níveis diferentes de “dimensionalidade”.

Os descritores 0D são uma representação mais simples e fragmentada da estrutura molecular, não se atentando a qualquer conhecimento sobre a conectividade dos átomos. São utilizados, por exemplo, a contagem de átomos e somas das propriedades atômicas; os descritores unidimensionais (1D) estão relacionados com as propriedades físico-químicas e fórmula molecular. Por exemplo, os tipos de átomos e grupos funcionais, efeitos estéricos,

hidrofóbicos e eletrônicos; os descritores bidimensionais (2D) consideram a conectividade dos átomos e compreendem descritores topológicos, constitucionais, geométricos e químicos-quânticos. Já os descritores tridimensionais (3D) capturam informações sobre a natureza e conectividade dos átomos e a configuração espacial geral da molécula no espaço 3D (TODESCHINI; CONSONNI; MANNHOLD, 2000). De forma ainda mais descritiva, a análise QSAR inclui métodos com mais dimensões, tais como o QSAR 4D (HOPFINGER *et al.*, 1997), 5D (VEDANI; DOBLER, 2002), 6D (VEDANI; BOBLER; LILL, 2005) e 7D (POLANSKI, 2009).

Em geral, a modelagem QSAR envolve um processo sistemático com várias etapas para a construção de um modelo confiável e preditivo. Geralmente, inclui-se a seleção do grupo de moléculas; geração de descritores; seleção dos descritores; construção do modelo e validação do modelo. Durante as primeiras etapas, na preparação do conjunto de dados, é importante que se atente à qualidade do conjunto de dados, para a construção de um modelo confiável. A segunda e terceira etapas, que consistem na seleção e geração dos descritores moleculares, são de suma relevância. A seleção de descritores que se correlacionam adequadamente com a atividade biológica é imprescindível para garantir a robustez da modelagem QSAR, para que capture as variações estruturais e informações geradas (YOUSEFINEJAD; HEMMATEENEJAD, 2015).

Após a seleção e geração dos descritores, a próxima etapa é a construção do modelo, a partir de métodos estatísticos adequados para encontrar a relação entre descritores e atividades biológicas. Por exemplo, pode-se utilizar abordagens lineares multivariadas, como a regressão por mínimos quadrados parciais (PLS) ou regressão linear múltipla (MLR), que são ferramentas quiomiométricas. Após a construção do modelo, é necessário validá-lo para verificar o poder de predição e estabilidade do modelo; para isso, utilizam-se de instrumentos estatísticos (YOUSEFINEJAD; HEMMATEENEJAD, 2015).

4.5 Multivariate Image Analysis applied to QSAR (MIA-QSAR)

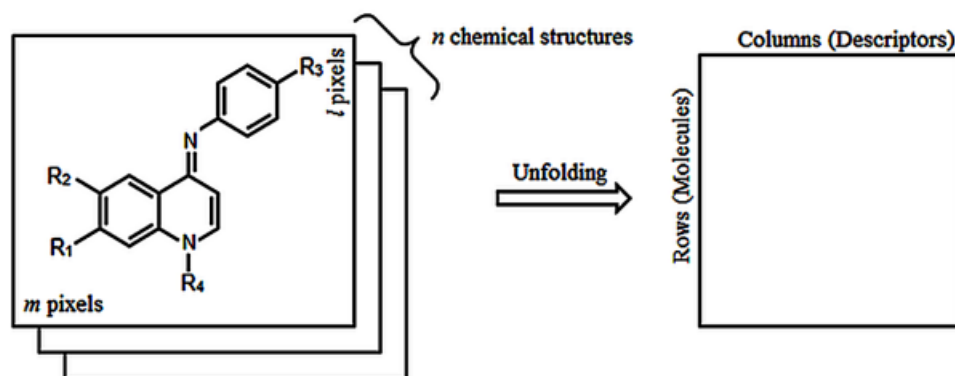
MIA-QSAR, ou mais especificamente, análise multivariada de imagens aplicada nas relações quantitativas entre estrutura e atividade (*Multivariate Image Analysis applied to Quantitative Structure-activity Relationships*) é uma abordagem QSAR bidimensional, que correlaciona desenhos de estruturas moleculares com a respectiva atividade biológica e/ou propriedades físico-químicas de compostos químicos. (CORMANICH; NUNES; FREITAS,

2012). Existem inúmeras vantagens na utilização dessa técnica, como ser um método acessível, o qual utiliza procedimentos e recursos computacionais de fácil execução, apresenta baixo custo, além de ser igualmente preditivo a métodos já bem consolidados, tais como CoMFA e CoMSIA (FREITAS; BROWN; MARTINS, 2005; FREITAS, 2006).

As imagens de estruturas moleculares podem trazer informações relevantes sobre o comportamento químico e físico de uma molécula, podendo ser úteis na descrição de propriedades topoquímicas e topoestruturais (informações sobre a forma da molécula, efeitos estéricos, etc). Desse modo, a abordagem MIA-QSAR parte do pressuposto de que alterações na estrutura química de um composto, ou a modificação de substituintes alterando as posições em uma série congênere de moléculas, resultam em alterações nas coordenadas dos pixels da imagem. Nessa perspectiva, o método baseia-se na utilização de desenhos de estruturas químicas desenhadas em programas específicos para esse fim, utilizando pixels (valores que correspondem a menor porção de uma imagem digital) de imagens como descritores moleculares (FREITAS, 2006; CORMANICH; NUNES; FREITAS, 2012).

O método MIA-QSAR precursor, aqui nomeado de MIA-QSARt (MIA-QSAR tradicional), é baseado na construção de imagens onde utiliza-se o sistema binário de pixels; os pixels de cor branca correspondem ao algarismo 765, enquanto os pixels pretos equivalem ao algarismo 0, de acordo com a escala RGB (*red; green; blue*). O branco é a soma dos canais vermelho, verde e azul, sendo 255 o valor para cada canal. Por sua vez, a cor preta é a ausência de cores. Desse modo, o número 765 representa a parte em branco na imagem, enquanto o 0 representa os átomos e ligações químicas presentes na molécula, ou seja, os *wireframes* (FREITAS; BROWN; MARTINS, 2005). A figura 5 a seguir representa um desenho de uma estrutura química e um *workflow* para modelagem segundo a abordagem MIA-QSARt.

Figura 5- Esquema utilizado para o tratamento das imagens e geração dos descritores MIA QSAR.

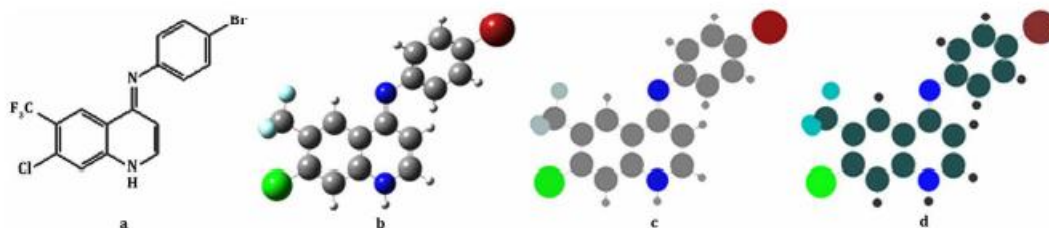


Fonte: BARIGYE; FREITAS (2016).

O sucesso dessa técnica pode ser averiguado em diversos trabalhos publicados na literatura (FREITAS; BROWN; MARTINS, 2005; GOODARZI; FREITAS; FERREIRA, 2009; CORMANICH; FREITAS; RITTNER, 2011). Entretanto, embora os resultados da primeira versão do MIA-QSAR fossem satisfatórios em diferentes análises e aplicações, encontrava-se alguns desafios e limitações em função do alinhamento impreciso dos *wireframes*. Geravam-se variações nas posições dos pixels devido ao desalinhamento, que, por sua vez, sugestionava o modelo de regressão, isto é, as insuficiências causadas na etapa de alinhamento tornavam-se a causa de erro sistemático para a análise. (GOODARZI; FREITAS; FERREIRA; 2009). Além disso, perdiam-se algumas informações contidas na estrutura química, devido ao tipo de representação, sistema binário de pixels, tais como o raio atômico e a natureza dos substituintes.

Desse modo, desenvolveu-se uma nova versão denominada *aug*-MIA-QSAR (*argmented* MIA-QSAR), a fim de superar as dificuldades encontradas no modelo anterior (NUNES; FREITAS, 2013; BARIGYE; FREITAS, 2016). Nessa abordagem, os átomos são retratados por círculos com dimensões proporcionais aos respectivos raios de van der Waals, e há a inclusão de cores que, de acordo com o sistema de cores RGB, variam de 0 (píxeis pretos) e 765 (píxeis brancos). Com a finalidade de esclarecimento, a Figura 6 mostra uma representação dos compostos no modelo *aug*-MIA-QSAR.

Figura 6- Imagens químicas empregadas para gerar os descritores MIA-QSAR de acordo com a) o MIA-QSAR tradicional, b) *aug*-MIA-QSARvol, c) *aug*-MIA-QSARvol (átomos com cores sólidas), e d) *aug*-MIA-QSARcolor (cores com valores de pixel proporcionais a propriedades atômicas).



Fonte: BARIGYE; FREITA (2016).

A incorporação desses aspectos melhorou muito a predição do modelo e, em 2016, a técnica passou por modificações que melhoraram ainda mais a sua versatilidade (BARIGYE; FREITAS, 2016). Essa modificação proporcionou a interpretação química dos modelos, pois cada átomo pode ser colorido de tal forma que o seu valor numérico seja proporcional a uma propriedade atômica, tal como o raio atômico e a eletronegatividade. Dessa forma, os

descritores MIA podem representar efeitos que determinam o modo de interação entre uma enzima e um ligante, tais como o tamanho (que se relaciona com interações estéricas e hidrofóbicas, por exemplo) e a eletronegatividade atômica (que se relaciona com interações dipolares, por exemplo).

No que tange às mudanças do método MIA-QSAR, outros melhoramentos foram introduzidos à técnica ainda em 2016. Desenvolveu-se uma ferramenta gráfica, o MIA-plot, que possibilita uma maior compreensão sobre a interpretação química dos modelos elaborados, isto é, o recurso criado ilustra como os diferentes substituintes e suas posições influenciam a propriedade de interesse. A técnica em questão consiste praticamente de dois gráficos coloridos, dois quais o primeiro é denominado de gráfico de coeficientes de regressão do PLS (**b**), que indica como (aumentando ou diminuindo) determinadas regiões na molécula afetam a variável resposta, enquanto o segundo é um gráfico de projeção da relevância das variáveis para o conjunto de compostos **VIP** (*Variable Importance in Projection*), que indica quais regiões na molécula mais afetam a variável resposta (BARIGYE *et al.*, 2016; BORGES; BARIGYE; FREITAS, 2017).

Formalmente, é recomendada uma análise conjunta de **b** e **VIP**, visto que um *score* **VIP** exhibe a relevância dos átomos nas moléculas para a explicação do modelo de projeção PLS, tanto para as variáveis preditoras, quanto para a variável resposta, permitindo, assim, visualizar o significado que as variáveis trazem para o modelo. Por outro lado, um valor significativo de **b**, seja um valor alto positivo ou negativo, infere que tal variável tem uma maior correlação com a propriedade modelada, ou seja, a variável afeta positivamente ou negativamente a resposta (BARIGYE *et al.*, 2016).

O método apresentado e descrito anteriormente, assim como outras abordagens QSAR, apresentam vantagens e desvantagens na compreensão da correlação entre estrutura-atividade. Assim sendo, compete ao pesquisador escolher a técnica que mais lhe é interessante para alcançar os objetivos de interesse. Entretanto, por mais que existam diferentes abordagens QSAR, existe um consenso de etapas, regras e critérios comuns, como já mencionado anteriormente, a serem cumpridas no desenvolvimento de um modelo robusto e realístico (TROPSHA, 2010).

A geração de descritores é uma delas, na qual correlacionam-se os pixels das imagens de estruturas químicas com a atividade biológica de um composto químico. Entretanto, o sucesso do modelo é dependente da representação correta da estrutura química, para que os descritores sejam fiéis às propriedades químicas e suas respectivas atividades. Após a geração, necessita-se selecionar os descritores; existem inúmeras formas de se extrair descritores de

imagens. No caso do MIA-QSAR, as informações de interesse são relacionadas com modificações estruturais que ocasionam alterações na propriedade biológica analisada. Desse modo, selecionam-se os descritores relevantes, excluindo os que não apresentam variações nas imagens, ou seja, as partes em branco das imagens, bem como a parte em comum de todas elas. Para isso, utiliza-se um método de regressão adequado para a análise multivariada do conjunto de dados.

4.6 Análise multivariada

Com o objetivo de eleger descritores para melhor se correlacionar com a atividade biológica nos modelos QSAR, procedimentos de seleção de variáveis podem ser fundamentais para excluir dados moleculares não fundamentais para descrever a atividade biológica, ou com baixa variância, bem como, com alta correlação. Para esse fim, o algoritmo genético (GA) é um método amplamente utilizado em QSAR, consiste em uma técnica de análise de dados, no qual otimiza as variáveis, baseada no princípio Darwiniano da evolução das espécies e na genética, no qual os métodos de cruzamento, seleção natural e mutação genética são capazes de viabilizar uma gama de populações mais evoluídas (FERREIRA; MONTANARI; GAUDIO, 2002). O método é útil para a otimização de dados para resolução de problemas envolvendo relações, capacidade de planejamento e tomada de decisões e seleção de subconjunto de dados, entre outros.

No desenvolvimento do GA ocorre as seguintes etapas: iniciação, seleção, cruzamento e mutação. Na primeira etapa, iniciação, as variáveis que pertencem ao grupo de dados denominados de população são agrupadas aleatoriamente, gerando os cromossomos. O agrupamento é realizado e definido por algum método de regressão multivariada e o valor de R (coeficiente de correlação) e RMSECV (raiz de erro quadrado médio da validação cruzada) são utilizados para estabelecer os melhores dados, denominados indivíduos essa etapa faz parte da seleção. No estágio de cruzamento, as variáveis (genes) dos indivíduos são mescladas na busca dos modelos melhores. Em seguida, mutações são realizadas para que todos as variáveis possam participar do processo, para evitar erros que possam ter ocorrido na seleção. O processo é finalizado após o preestabelecimento do número de gerações no momento em que apresentar uma porcentagem de indivíduos na população estarem usando variáveis idênticas. O modelo gerado, usualmente é avaliado por parâmetros estatísticos: (R), o fator F e o nível de preditividade (Q²) (FERREIRA; MONTANARI; GAUDIO, 2002; TEÓFILO, 2013).

Um dos passos cruciais em QSAR consiste na geração de um modelo matemático, no qual ocorre a correlação da atividade biológica (**Y**) e os descritores (**X**), por meio de, na maior parte, equações de regressão linear. Em geral, são utilizados métodos de calibração multivariada pelo grande número de variáveis descritivas (variáveis independentes); no caso do MIA-QSAR, esse número pode chegar a dezenas de milhares de variáveis a (FERREIRA *et al.*, 1999; DAMALE *et al.*, 2014).

Nos últimos anos, vários métodos quimiométricos ampliaram as ferramentas que podem ser utilizadas nos estudos QSAR e, dentre eles, os mais usados são: os métodos quantitativos de Regressão Linear Múltipla (MLR) e Mínimos Quadrados Parciais (PLS) (MARTINS; FERREIRA, 2013).

4.6.1 MLR - Regressão Linear Múltipla

MLR (do inglês *Multiple Linear Regression*) é, historicamente, o primeiro método de calibração multivariada usado em modelos QSAR. Uma das principais vantagens da MLR é a sua simplicidade, tornando-se uma técnica extensamente utilizada. A sua forma geral é apresentada a seguir:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + e \quad (2)$$

Em que, b_1 até b_n são coeficientes de regressão, b_0 é a constante. A partir do modelo gerado pode-se tanto investigar como cada descritor influencia a variável resposta, quanto avaliar as implicações conjuntas dos descritores na variável resposta (LIU; LONG, 2009). Uma das desvantagens da utilização da MLR é que possui limitações em relação à colinearidade dos descritores, o que pode gerar um modelo impreciso. Além disso, existe a condição de a quantidade de descritores não exceder o número de amostras do conjunto de treinamento, tornando-se uma problemática ao se lidar com conjuntos químicos, pois costuma-se ter um número bem maior de variáveis independentes comparado ao de amostras (WOLD; SJÖSTRÖM; ERIKSSON, 2001; LIU; LONG, 2009). Ao gerar-se um modelo com muitos descritores, ocorre um problema denominado *overfitting*, isto é, ocorre um superajuste do modelo.

4.6.2 PLS - Mínimos Quadrados Parciais

A técnica dos mínimos quadrados parciais, PLS (do inglês *Partial Least Squares*), ao contrário da técnica MLR, consegue analisar dados fortemente colineares e com inúmeras variáveis **X**. Além disso, o método é mais robusto, pois permite avaliar os dados de maneira mais realista, por passar por uma etapa que assegura que suas componentes principais

produzidas sejam relevantes para a atividade biológica. Analisam-se os conjuntos \mathbf{X} e \mathbf{y} simultaneamente, baseando-se no método de covariância, ou seja, é iterativo, analisando os erros em \mathbf{y} à medida que analisa o \mathbf{X} . Assim, as componentes principais são otimizadas com a finalidade de melhor descrever a relação entre \mathbf{X} e \mathbf{Y} (por esse motivo, são descritas como variáveis latentes) (FERREIRA; MONTANARI; GAUDIO, 2002).

Em PLS, a matriz \mathbf{T} contém as variáveis latentes das amostras, enquanto a que contém as variáveis com informações acerca da resposta é chamada de \mathbf{U} . As variáveis latentes \mathbf{T} (*X-scores*) são usadas para predizer \mathbf{U} (*Y-scores*), e esses são empregados para construir previsões para as respostas. Ressalta-se que os \mathbf{X} e *Y-scores* são escolhidos de forma que a correlação entre pares sucessivos de *scores* seja a maior possível (TOBIAS, 2016).

Além desses, existem outros métodos de calibração multivariada lineares, como o PCR, e não-lineares (algoritmos genéticos, redes neurais, etc.). O desempenho desses algoritmos e de outros é avaliado por meio de parâmetros de validação. Nessa etapa, permite-se verificar a acurácia, a estabilidade, bem como a exatidão da previsão da variável de interesse.

4.6.3 Validação QSAR

A etapa de validação dos modelos QSAR é, sem dúvida, uma das principais e mais importante, pois a partir dela é que se mede a confiabilidade do modelo desenvolvido (YOUSEFINEJAD; HEMMATEENEJAD, 2015). A validação do modelo QSAR é efetuada e fundamentada em parâmetros de validação interna e externa. Com isso, existe a separação do conjunto de dados em grupo teste e treinamento. Os dados do grupo de treinamento são utilizados, basicamente, para a calibração do modelo, enquanto os dados do grupo teste são utilizados para validar o modelo (TROPISHA, 2010; ROY et al., 2012).

A validação interna, em PLS, é responsável por determinar o número de variáveis latentes (LV), bem como identificar *outliers*. Quanto a validação externa, ela é exclusiva para verificar quão preditivo é o modelo e a sua confiabilidade. Para realizar a validação do modelo QSAR, é necessário utilizar parâmetros estatísticos confiáveis e que permitam verificar a acurácia do modelo.

Desse modo, existem inúmeros parâmetros que podem ser utilizados, os mais comuns são: r^2 (coeficiente linear de determinação), q^2 (coeficiente de determinação da validação interna cruzada), r^2_{test} (coeficiente de determinação da validação externa), $r^2_{\text{y-rand}}$ (coeficiente de determinação obtido após a randomização do bloco das variáveis dependentes), RMSE (erro quadrático médio), dentre outros.

A divisão do grupo teste e grupo treinamento pode ser realizada seguindo a forma randômica, ou por outro método estatístico. Uma técnica estatística de amostragem de grande utilização é o algoritmo Kennard-Stone (KENNARD; STONE, 1969). Esse algoritmo tem por finalidade selecionar um subconjunto de amostras \mathbf{X} (grupo teste), com base em um conjunto mais abrangente de amostras \mathbf{Y} (grupo treinamento); com isso, as amostras do subconjunto \mathbf{X} representam o máximo de variabilidade do conjunto \mathbf{Y} . O algoritmo seleciona as amostras com base na distância euclidiana, para assegurar uma distribuição uniforme do subconjunto.

A validação cruzada *leave-one-out* (LOOCV) é um tipo de validação interna bastante comum. Ela consiste na separação dos n dados em duas partes: retira-se uma amostra do conjunto de treinamento, isto é, amostra teste, para a fase da validação, o restante das amostras é utilizado para a calibração e, por fim, as medidas da regressão obtidas são utilizadas para previsão da atividade biológica da amostra deixada de fora da calibração. Repete-se esse procedimento n vezes, excluindo em cada momento uma observação diferente (ROY, et al, 2012). Como resultado, obtém-se o coeficiente de determinação da validação cruzada, r^2_{cv} (q^2), cujos valores acima de 0,5 são um critério de robustez e com boa capacidade preditiva (GOLBRAIKH; TROPSHA, 2002).

Para avaliar a capacidade de modelagem dos descritores, consideram-se as métricas de erro quadrático médio da calibração (RMSEc) e o coeficiente de determinação (r^2), calculado para o conjunto treinamento. Já os resíduos, isto é, a diferença entre os valores experimentais e estimados do modelo, são utilizados para calcular o erro quadrado médio na validação cruzada (RMSEcv). Consideram-se as estatísticas de modo que $r^2 > q^2$ e $RMSEcv > RMSEc$. Em QSAR, admite-se um valor mínimo de $r^2 > 0,6$ (KIRALJ; FERREIRA, 2009).

Um outro método de validação é o *y-randomization*, cujo intuito é verificar as correlações ao acaso entre as variáveis dependente e independente. Desse modo, a técnica busca encontrar descritores que estabelecem boa relação com a variável resposta (\mathbf{y}). O teste de randomização de \mathbf{y} consiste em inúmeras execuções que se utiliza a matriz de variáveis independentes originais (\mathbf{X}) fixa e apenas o bloco das variáveis \mathbf{y} é aleatorizado (KIRALJ; FERREIRA, 2009). Após a randomização de \mathbf{y} , o coeficiente de determinação (r^2_{rand}) é obtido e essa medida é utilizada para averiguar se o r^2_{cal} e o r^2_{cv} não foram ajustados ao acaso. Os modelos adquiridos nas condições descritas anteriormente devem ser consideravelmente ruins, esperando-se que os valores de r^2_{rand} sejam bem menores que r^2_{cal} (RÜCKER; RÜCKER; MERINGER, 2007).

Para a certificação de que r_{rand}^2 seja substancialmente menor que r_{cal}^2 e verificar a confiabilidade do modelo, utiliza-se o coeficiente de determinação penalizado e corrigido ${}^c r_p^2$ (Equação 3), verificando a relevância entre as diferenças dos coeficientes r_{cal}^2 e r_{rand}^2 .

$${}^c r_p^2 = r \times (r_{\text{cal}}^2 - r_{\text{rand}}^2)^{1/2} \quad (3)$$

Além das validações internas, necessita-se validar o modelo externamente para testar quão preditivo é. Para isso, utilizam-se de parâmetros como o coeficiente de determinação modificado r_m^2 , que verifica a similaridade entre os valores experimentais e preditos, além, é claro, do r_{test}^2 . Para que o modelo possa ser considerado preditivo, ambos os coeficientes devem apresentar valores superiores a 0,5. Outros parâmetros estatísticos são usados, como o erro quadrado médio de predição (RMSEp), o coeficiente de correlação externa (Q_{ext}^2), dentre outros. Q_{ext}^2 é análogo ao q^2 apresentado anteriormente.

Percebe-se que, não é apenas um parâmetro que avalia a previsibilidade e robustez de um modelo QSAR, mas um conjunto de fatores que auxiliam para essa interpretação.

5 DOCKING MOLECULAR

Docking molecular ou ancoramento molecular consiste em uma abordagem de modelagem molecular que facilita e otimiza o desenvolvimento de novos compostos bioativos (SANT'ANNA, 2009). Muitas vezes, é utilizado com associação à técnica QSAR para complementar o desenvolvimento de bioligantes (por exemplo, herbicidas e fármacos). Como visto anteriormente, a modelagem QSAR é baseada na estrutura do ligante, isto é, um método independente do receptor. Já o *docking* molecular resume-se na predição da conformação bioativa do ligante no sítio ativo da macromolécula e é uma das principais estratégias do método SBDD (do inglês, *structure-based drug design*). Em geral, os métodos SBDD são estruturados no arranjo topológico de alvos biológicos (SANT'ANNA, 2009).

Além da predição do posicionamento do ligante no sítio de ligação de um receptor, o procedimento de *docking* permite verificar a complementaridade estrutural entre ligante-macromolécula, o que proporciona estimar, classificar e identificar as interações intermoleculares específicas e mais favoráveis que regem a ligação proteína-ligante (DU et al., 2016). Isso ocorre devido ao acoplamento molecular do ligante otimizado dentro do receptor, assim, a menor energia é atribuída a uma certa pontuação, que é expressa como uma probabilidade de ligação (CAVASOTTO et al., 2018).

A avaliação da afinidade de cada uma das conformações geradas no ancoramento molecular é realizada por funções score. Os valores da função score, E_{score} , são definidos pela Equação 4:

$$E_{score} = E_{inter} + E_{intra} \quad (4)$$

em que o termo E_{inter} , que corresponde à energia de interação ligante-proteína, é definido de acordo com a Equação 5.

$$E_{inter} = \sum_{i \in \text{ligante}} \sum_{j \in \text{proteína}} \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4r_{ij}} \right] \quad (5)$$

O primeiro termo, E_{PLP} , representado na Equação 3, é um termo de energia potencial que utiliza dois conjuntos diferentes de parâmetros: um para a aproximação do termo estérico (van der Waals) entre átomos e outro potencial para a ligação hidrogênio. O segundo termo é responsável por descrever as interações eletrostáticas entre os átomos. O valor numérico de 332,0 fixa as unidades de energia eletrostática para kcal mol⁻¹ (THOMSEN; CHRISTENSEN, 2006).

O termo E_{intra} , que corresponde à energia interna do ligante, é definido de acordo com a Equação 6.

$$E_{intra} = \sum_{i \in \text{ligante}} \sum_{j \in \text{ligante}} E_{PLP}(r_{ij}) + \sum_{\text{ligações flexíveis}} A[1 - \cos(m \cdot \theta - \theta_0)] + E_{penalizada} \quad (6)$$

As duas primeiras somas referem-se a todos os pares de átomos do ligante, excluindo-se os pares de átomos conectados por duas ligações. O segundo termo refere-se à energia de torção, em que θ é o ângulo de torção da ligação. A média da contribuição de torção da ligação da energia é utilizada se diversas torções são determinadas. O último termo, $E_{penalizada}$ (termo de correção), atribui uma penalidade de 1.000 se a distância entre dois átomos pesados (mais de duas ligações distantes) for menor que 2,0 Å, punindo conformações inexistentes do ligante. Em resumo, essas funções são utilizadas para ancorar automaticamente uma molécula flexível em uma molécula molde (proteína) (THOMSEN; CHRISTENSEN, 2006).

6 MATERIAIS E MÉTODOS

Para a modelagem MIA-QSAR, necessita-se escolher um conjunto de dados, sendo este uma série de compostos congêneres. Portanto, na primeira parte do trabalho, selecionou-se um conjunto de dados de uma série de pirimidil acil tiureias (LI et al. 2021) Tabela 1 e Figura 3.

Tabela 1- 14 N-(4-metanossulfonil)benzoil-(N'-pirimidin-2-il)tiureias usados na modelagem MIA-QSAR e a respectiva inibição experimental do crescimento radicular do crescimento da *B. napus* (% , em decimal) e milog P calculando usando o programa Molinspiration (www.molinspiration.com).

Compostos	R ¹	R ²	Inibição	Milog P
1	OMe	OMe	0,5527	1,28
2	OEt	OEt	0,4558	2,04
3	OiPr	OiPr	0,4363	2,76
4	Cl	OMe	0,6913	1,91
5	Cl	OEt	0,4596	2,28
9	Cl	Cl	0,6240	2,53
7	Cl	NEt ₂	0,5907	2,70
8	OMe	NEt ₂	0,6001	2,08
9	OEt	NEt ₂	0,5224	2,46
10	OiPr	NEt ₂	0,5109	2,82
11	Me	Cl	0,6359	1,95
12	Me	OMe	0,5227	1,32
13	Me	OEt	0,3384	1,70
14	OMe	OH	0,5660	1,01

Fonte: Do autor (2022).

Desse modo, a análise QSAR empregou 14 N-(4-metanossulfonil)benzoil-(N'-pirimidin-2-il)tiureias substituídas, previamente sintetizadas e testadas contra *B. napus L.*, cujas atividades herbicidas são descritas em termos de inibição média do crescimento radicular (%) mediante aplicação de soluções nas concentrações de 1, 10 e 100 mg L⁻¹ (LI et al. 2021). Os descritores QSAR aplicados como preditores no modelo de mínimos quadrados parciais (PLS) são baseados em análise multivariada de imagens, correspondendo a unidades digitais (pixels) de imagens químicas 2D que representam as estruturas moleculares.

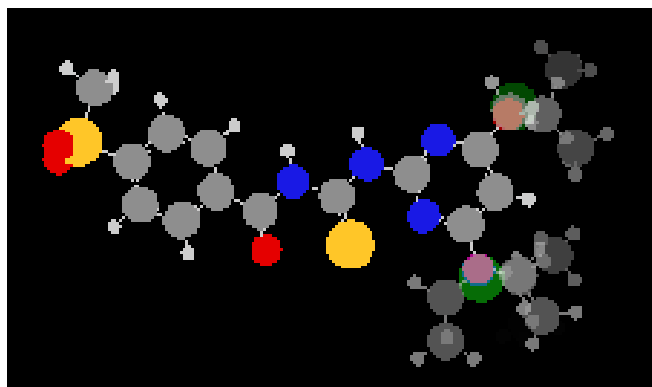
Essas imagens foram esboçadas usando o programa GaussView (DENNIGTON et al. 2008) e são sobreponíveis pela subestrutura congênere correspondente à porção 4-*N*-(metanossulfonil)benzoil-(*N'*-pirimidin-2-il)tiureia (Figura 7). Os átomos nas moléculas foram representados por círculos com tamanhos proporcionais aos raios de van der Waals, para codificar propriedades estéricas e hidrofóbicas. Além disso, considerou-se que as cores dos pixels correspondem aos valores RGB (*red-green-blue*); a cor para cada tipo de átomo foi numerada proporcionalmente à eletronegatividade de Pauling. Cada imagem de dimensão 200 × 228 pixels foi desdobrada em um vetor linha e, em seguida, agrupada para formar uma matriz de descritores de tamanho 14 × (200 × 228) pixels.

O modelo de calibração foi obtido por regressão da matriz **X** contra o bloco **y** contendo os dados de atividade herbicida. Para avaliar se o resultado da calibração foi obtido por correlação casual ou não, foi realizado um teste de randomização **y** (10 ciclos) e o coeficiente de determinação obtido após aleatorizar o bloco **y** - r^2_{y-rand} - foi comparado com o r^2 verdadeiro usando o parâmetro ${}^c r^2_p$ (valores acima de 0,5 são aceitáveis) (MITRA et al. 2010). A ausência de outliers foi verificada através do gráfico de *leverages* × resíduos de Student (ROY et al. 2015). O modelo foi validado interna e externamente usando validação cruzada *leave-one-out* e um procedimento de *bootstrapping*, respectivamente (CONSONNI et al. 2009; KIRALJ e FERREIRA, 2009). A validação por *bootstrapping* consistiu em selecionar aleatoriamente, por 10 vezes, os compostos do conjunto de teste e, em seguida, obter o desempenho estatístico para cada ciclo. O desenvolvimento do QSAR foi inteiramente realizado utilizando o programa Chemoface (NUNES et al. 2012). O parâmetro de lipofilicidade *milog P* foi calculado usando o programa Molinspiration (www.molinspiration.com).

A função de pontuação Moldock é um algoritmo de busca híbrido que combina a técnica de otimização de evolução diferencial com um algoritmo de previsão de cavidade durante o procedimento de busca, o que permite o reconhecimento rápido e preciso de potenciais modos de ligação, as poses. Ele é derivado do Potencial Linear por Partes (PLP), um potencial simplificado cujos parâmetros são adequados às estruturas de ligantes de proteínas e funções de pontuação de dados de ligação e estendidos no programa GEMDOCK (Método Evolutivo Genérico para DOCK molecular) com um novo termo de ligação de hidrogênio. O MolDock SE foi usado como algoritmo de busca, e o número máximo de poses geradas foi 100. Os ligantes e as cadeias laterais de aminoácidos foram flexíveis durante o encaixe. Os potenciais locais de ligação foram identificados usando o algoritmo de detecção de cavidade integrado. Uma grade cobrindo a proteína é criada e em cada ponto da grade, uma esfera é colocada. Verifica-se se esta esfera irá se sobrepor a alguma das esferas determinadas pelos raios de Van

der Waals dos átomos da proteína. As cavidades que são identificadas pelo algoritmo de detecção de cavidade são, então, usadas pelo algoritmo de busca de evolução diferencial guiada para focar a busca naquela área específica durante a simulação de encaixe.

Figura 7- Superposição de imagens correspondentes às 14 4-*N*-(metanossulfonil)benzoil-(*N'*-pirimidin-2-il)tiureias utilizadas na modelagem MIA-QSAR. As moléculas no sistema bola e bastão foram esboçadas no programa GaussView (Dennington et al. 2008) e os tamanhos dos átomos são proporcionais aos raios de van der Waals correspondentes. As cores dos átomos (valores de pixel RGB) foram renumeradas para fazer correspondência com a eletronegatividade de Pauling. Observe que as porções congruentes estão perfeitamente alinhadas, enquanto os substituintes variáveis nas posições R¹ e R² da estrutura vista na Figura 3 explicam a variância no bloco *y*.



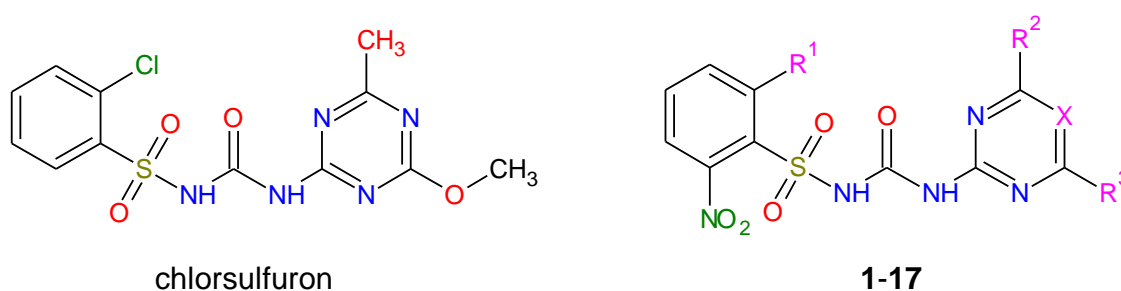
Fonte: Do autor (2022).

Um estudo de encaixe molecular foi realizado para explicar a afinidade de ligação das 14 *N*-(4-metanossulfonil)benzoil- *N'*-(pirimidin-2-il)tiureias contra a proteína PSII (photosystem II) (EC:1.10.3.9). Essa estrutura exibe uma rica coleção de detalhes atômicos: 16 clorofilas, 4 feofitinas, 8 betacarotenos, 2 plastoquinonas, 2 protoporfirinas IX contendo Fe, 2 aglomerados de Ca-Mn₄-O₅, 4 íons cloreto, 2 íons Fe(II) e 2 íons bicarbonato. A estrutura cristalina de *B. napus* PSII (*Bn* PSII) foi obtida de uma modelagem por homologia (FISER, 2010). A modelagem começou com uma busca sistemática no Swiss Bioinformatics Resource Portal (Expasy) para obter a estrutura primária do *Bn*PSII usando o banco de dados UniProt (The UniProt Consortium, 2021) como ferramenta de busca. Esse perfil foi então usado para pesquisar no PDB estruturas de proteínas conhecidas usando a sequência alvo como consulta. A sequência primária da proteína alvo foi alinhada com a proteína molde usando o programa de similaridade SIM (HUANG e MILLER, 1991). Os templates foram classificados com base na pontuação de alinhamento e na qualidade estrutural de acordo com WHAT_CHECK

(HOOFT, et al., 1996). O *BnPSII* foi construído usando o servidor SWISS-MODEL (WATERHOUSE et al., 2018). Os ligantes preparados foram encaixados no sítio de ligação da *BnPSII* usando o software Molegro Virtual Docker (THOMSEN e CHRISTENSEN, 2006), que é projetado para a previsão da conformação de ligação mais provável de um ligante em direção a uma macromolécula (RAMALHO et al., 2010).

Para a segunda parte do trabalho, uma série de 17 sulfonilureias (**10a –10q**) da Figura 8 e Tabela 2 foi recuperada da literatura (WEI; CHENG; et al., 2017), cujas taxas de inibição foram obtidas em três concentrações (0,1, 1,0 e 10 mgL⁻¹) e expressas em porcentagem de *crescimento de B. napus*. Essas concentrações foram convertidas para molL⁻¹ e depois para pIC₅₀ usando o método probit (FINNEY, 1952). Esses dados foram regredidos contra descritores MIA selecionados por meio de algoritmo genético usando regressão linear múltipla. Os descritores MIA foram gerados a partir de imagens moleculares esboçadas no programa GaussView (DENNINGTON; et al. 2008), onde os átomos são representados como círculos coloridos com tamanhos proporcionais aos respectivos raios de van der Waals. Como as cores são descritas pelo sistema RGB, cujos valores numéricos variam de zero (preto) a 765 (branco), os pixels de cada tipo de átomo foram renumerados para fazer uma correspondência com o respectivo raio de van der Waals (r_{vdw}) e eletronegatividade (ϵ) valores.

Figura 8- Clorsulfuron e sulfonilureias aqui estudadas.



Fonte: Do autor (2022).

Cada imagem de dimensão 300x354 pixels foi desdobrada em um vetor linha de 106.200 pixels e agrupada para obter uma matriz de dados de 17 × 106.200. Os descritores MIA mais relevantes foram selecionados por meio de algoritmo genético usando uma função de aptidão baseada em parâmetros fundamentados em erro/MAE e 100 iterações/gerações, probabilidade de mutação de 0,3 e 30 equações selecionadas em cada geração. Como resultado, cinco descritores selecionados para cada tipo de descritor MIA ($MIA_{r_{vdw}}$ e MIA_{ϵ}) foram regredidos contra os valores de pIC₅₀ usando regressão linear múltipla (MLR) para obter os

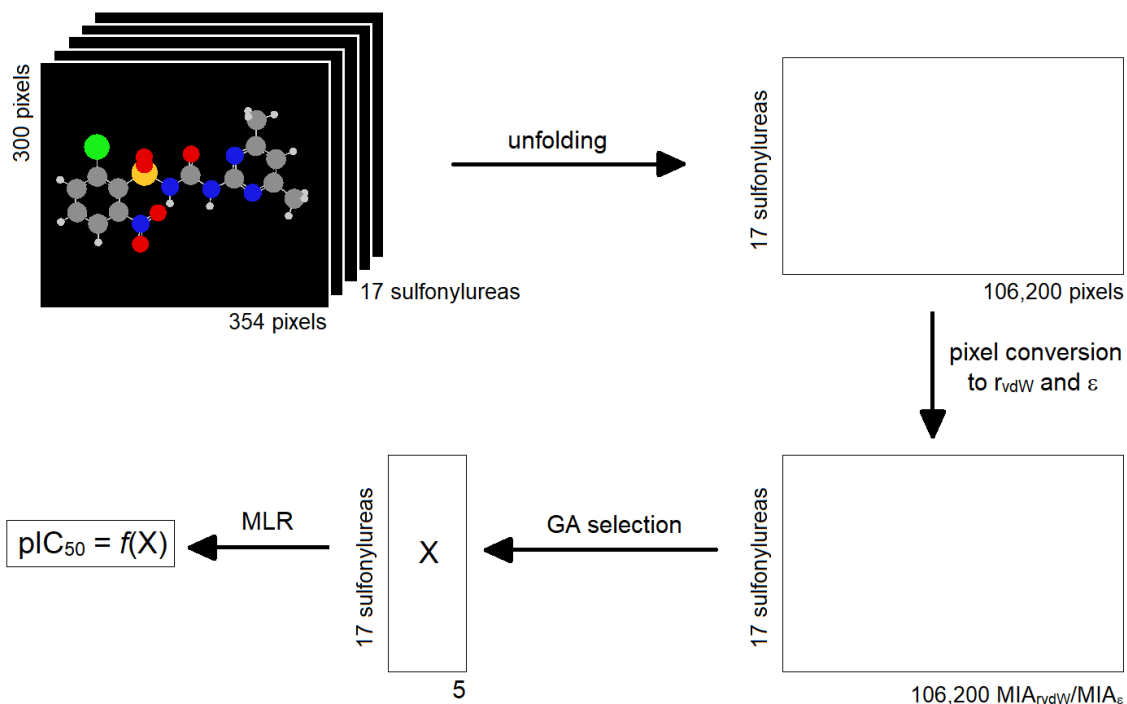
modelos MIA-QSAR (Figura 9). Para atestar o desempenho preditivo desses modelos, as matrizes **X** de dimensão 17×5 foram divididas em conjuntos de treinamento (13 amostras) e teste (4 amostras) usando Kennard-Stone (KENNARD, 1969) e amostragem aleatória. Os modelos resultantes foram validados de acordo com as melhores práticas em QSAR (TROPISHA, 2010). As matrizes de descritores originais também foram regredidas contra os valores pIC_{50} usando regressão de mínimos quadrados parciais (PLS) para gerar importância variável nas pontuações de projeção (VIP) e coeficientes de regressão PLS (**b**), que foram aplicados para construir MIA-plots para entender como os substituintes afetam as variáveis de resposta. As pontuações VIP descrevem o efeito de cada variável no modelo PLS e são calculadas como a soma ponderada dos quadrados dos pesos, que mede a porcentagem do **y** variância explicada com cada variável latente. Por sua vez, **b** corresponde ao vetor de coeficientes de relação interna PLS e dá a relevância de cada variável original para **y**. Este procedimento foi realizado com o programa Chemoface.

Tabela 2- Série das sulfonilureias e respectivos valores de pIC_{50} (IC_{50} em mol L⁻¹).

Composto	X	R ¹	R ²	R ³	pIC_{50}
10 ^a	CH	Cl	CH ₃	CH ₃	6,35
10 ^b	CH	Cl	OC ₃	OCH ₃	8,92
10 ^c	CH	Cl	H	CH ₃	5,20
10 ^d	CH	OCH ₂ CF ₃	CH ₃	CH ₃	8,65
10 ^e	CH	Cl	CH ₃	OCH ₃	5,41
10 ^f	CH	Cl	Cl	OCH ₃	5,83
10 ^g	N	Cl	CH ₃	OCH ₃	5,13
10 ^h	CH	F	OCH ₃	OCH ₃	6,26
10 ⁱ	CH	F	CH ₃	OCH ₃	6,54
10 ^j	CH	F	H	CH ₃	5,27
10 ^k	N	F	CH ₃	OCH ₃	4,25
10 ^l	CH	I	CH ₃	OCH ₃	6,35
10 ^m	CH	I	OCH ₃	OCH ₃	5,57
10 ⁿ	CH	I	H	CH ₃	5,11
10 ^o	CH	Br	CH ₃	OCH ₃	6,40
10 ^p	CH	Br	OCH ₃	OCH ₃	5,46
10 ^q	N	Br	CH	OCH ₃	5,22

Fonte: Do autor (2022).

Figura 9- Workflow seguido para obtenção dos modelos MIA-QSAR.



Fonte: Do autor (2022).

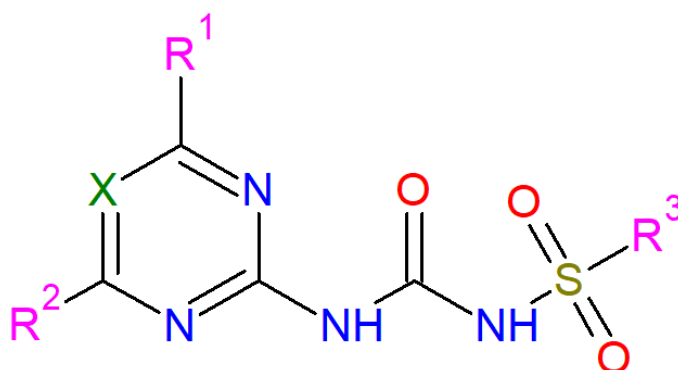
Um terceiro estudo foi realizado para verificar as características estruturais que afetam a sorção no solo de herbicidas comerciais conhecidos. Para isso, os valores de $\log K_{OC}$ para 16 sulfonilureias medidos pelo *Relatório Nacional sobre Exposição Humana a Produtos Químicos Ambientais* foram obtidos do banco de dados *PubChem* e foram relatados principalmente pela Toxicidade Pesquisável de Estrutura Distribuída da EPA dos EUA (Tabela 3 e Figura 10).

O pH padrão 7 a 20°C está implícito. Os descritores moleculares utilizados para análise da relação estrutura-atividade (SAR) foram baseados na análise multivariada de imagens (MIA), cujo procedimento de modelagem é descrito a seguir. As estruturas moleculares foram geradas usando o programa GaussView (DENNINGTON, et al., 2008), onde os átomos são representados como círculos coloridos com tamanhos proporcionais aos respectivos raios de van der Waals. As estruturas químicas 2D superponíveis foram convertidas em imagens bitmap de 300×354 pixels (Figura 10). Os valores dos pixels foram renumerados para fazer uma correspondência com a eletronegatividade de Pauling (modelo MIA- ϵ), raio de van der Waals (modelo MIA- r_{vdW}) e razão r_{vdW}/ϵ (modelo MIA- r_{vdW}/ϵ). As imagens foram unidas para formar uma matriz de três vias de dimensão $16 \times 300 \times 354$, que foi então desdobrada em uma matriz de $16 \times (300 \times 354)$. Essa matriz X foi submetida à análise de componentes principais (PCA) sem qualquer pré-processamento de dados usando o programa Chemoface (NUNES; et

al., 2012). Além disso, a matriz **X** foi regredida contra os valores de $\log K_{OC}$ usando regressão de mínimos quadrados parciais (PLS) e 2 variáveis latentes para obter os coeficientes de regressão PLS (**b**) e importância variável nas pontuações de projeção (VIP). Esses parâmetros foram usados para construir os gráficos MIA (BARIGYES; et al., 2016).

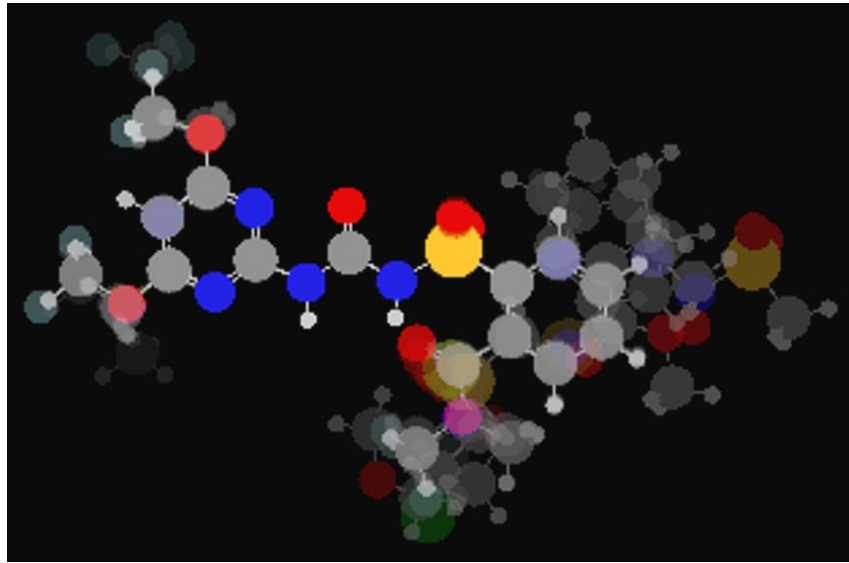
Um modelo MIA-QSAR usando descritores baseados em eletronegatividade selecionados por algoritmo genético (GA) foi construído, e esses descritores foram regredidos contra os valores de $\log K_{OC}$ usando regressão linear múltipla (MLR). A função de aptidão empregada na seleção do GA foi baseada em parâmetros baseados em erro/MAE, e um total de 100 iterações, probabilidade de mutação de 0,3 e 30 equações selecionadas em cada geração foram aplicadas (AMBURE et al., 2015). A qualidade do modelo foi determinada pelo coeficiente de determinação da calibração (r^2), validação cruzada leave-one-out (q^2) e predição (r^2_{pred}).

Figura 10- Estrutura básica das sulfonilureias estudadas neste trabalho.



Fonte: Do autor 2022.

Figura 11- Imagens sobrepostas dos 16 herbicidas sulfonilureias estudados. Observe as porções congruentes e variáveis (substituintes) responsáveis pelos dados de sorção do solo



Fonte: Do autor.

Tabela 3- Herbicidas sulfonilureias utilizados no estudo das propriedades estruturais das sorções no solo ($\log K_{oc}$).

Numero	Nome	X	R ¹	R ²	R ³	$\log K_{oc}$
1	Bensulfuron methyl	CH	OMe	OMe	-CH ₂ (2-COOMe)Ph	2,50
2	Foramsulfuron	CH	OMe	OMe	-(2-CONMe ₂ ,5-NHCHO)Ph	1,00
3	Mesosulfuron methyl	CH	OMe	OMe	-(2-COOMe,5-CH ₂ NHSO ₂ Me)Ph	1,04
4	Nicosulfuron	CH	OMe	OMe	-2(3-CONMe ₂)Pyr	1,48
5	Rimsulfuron	CH	OMe	OMe	-2(3-SO ₂ Et)Pyr	1,71
6	Sulfosulfuron	CH	OMe	OMe	-3(2-(SO ₂ Et)imidazo[1,2-a]pyridine)	1,52
7	Primisulfuron methyl	CH	OCF ₃	OCF ₃	-(2-COOMe)Ph	1,08
8	Sulfometuron methyl	CH	Me	Me	-(2-COOMe)Ph	1,62
9	Oxasulfuron	CH	Me	Me	-(2-COOC ₃ H ₆ O)Ph	1,93
10	Chlorsulfuron	N	OMe	Me	-(2-Cl)Ph	2,19
11	Metsulfuron methyl	N	OMe	Me	-(2-COOMe)Ph	1,54
12	Prosulfuron	N	OMe	Me	-(2-CH ₂ CH ₂ CF ₃)Ph	1,15
13	Triasulfuron	N	OMe	Me	-(2-OCH ₂ CH ₂ Cl)Ph	2,02
14	Thifensulfuron methyl	N	OMe	Me	-3(2-COOMe)thiophene	1,45
15	Ethametsulfuron methyl	N	OEt	NHMe	-(2-COOMe)Ph	1,93
16	Triflusulfuron methyl	N	OCH ₂ CF ₃	NMe ₂	-2-COOMe,6-Me)Ph	1,84

Fonte: Do autor (2022).

REFERÊNCIAS BIBLIOGRÁFICAS

AMBURE, P.; AHER, R. B.; GAJEWICZ, A.; PUZYN, T. *et al.* “NanoBRIDGES” software: Open access tools to perform QSAR and nano-QSAR modeling. **Chemometrics and intelligent laboratory systems**, 147, p. 1-13, 2015.

ARROIO, A.; HONÓRIO, K. M.; SILVA, A. B. F. D. Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade Quantum chemical properties used in structure-activity relationship studies. **Química nova**, 33, n. 3, p. 694-699, 2010.

BARIGYE, S. J.; DE FREITAS, M. P. Ten years of the MIA-QSAR strategy: historical development and applications. **International Journal of Quantitative Structure-Property Relationships (IJQSPR)**, 1, n. 1, p. 64-77, 2016.

_____; DUARTE, M. H.; NUNES, C. A.; FREITAS, M. P. MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. **RSC Advances**, 6, n. 55, p. 49604-49612, 2016. 10.1039/C6RA09593C.

BOLAÑOS-JIMÉNEZ, J.; USCANGA-MORTERA, E.; TAFOYA-RAZO, J. A.; KOHASHI-SHIBATA, J. *et al.* Efectividad biológica de herbicidas inhibidores de la acetil coenzima a carboxilasa y acetolactato sintasa y la presencia de resistencia en *Echinochloa crus-galli* (L.) Beauv. **Agrociencia**, 52, n. 5, p. 713-723, 2018-08 2018.

BORGES, C.; BARIGYE, S.; FREITAS, M. Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients. **Molecular Physics**, 115, p. 1-7, 07/05 2017.

BOSCHIN, G.; D’AGOSTINA, A.; ANTONIONI, C.; LOCATI, D. *et al.* Hydrolytic degradation of azimsulfuron, a sulfonylurea herbicide. **Chemosphere (Oxford)**, 68, n. 7, p. 1312-1317, 2007.

BRUSAMARELLO, A. P.; OLIVEIRA, P. H.; TREZZI, M. M.; XAVIER, E. *et al.* INHERITANCE OF RESISTANCE TO PROTOPORPHYRINOGEN OXIDASE INHIBITOR HERBICIDES IN WILD POINSETTIA. **Planta Daninha**, 34, n. 3, p. 575-580, 2016-09 2016.

CARVALHO, P. O. M. D.; FERREIRA, M. M. C. 2D, 3D and Hybrid QSAR Studies of Nostoclode Analogues as Inhibitors of the Photosystem II. **Journal of the Brazilian Chemical Society**, 30, n. 2, p. 265-278, 2019-02 2019.

CAVASOTTO, C. N.; AUCAR, M. G.; ADLER, N. S. Computational chemistry in drug lead discovery and design. **International journal of quantum chemistry**, 119, n. 2, p. e25678-n/a, 2019.

CECHIN, J.; VARGAS, L.; AGOSTINETTO, D.; LAMEGO, F. P. *et al.* Mutation of Trp-574-Leu ALS gene confers resistance of radish biotypes to iodosulfuron and imazethapyr herbicides. **Acta Scientiarum. Agronomy**, 39, n. 3, p. 299-306, 2017-09 2017.

_____; VARGAS, L.; AGOSTINETTO, D.; ZIMMER, V. *et al.* Resistance of Radish Biotypes to Iodosulfuron and Alternative Control. **Planta Daninha**, 34, n. 1, p. 151-160, 2016-03 2016.

CHEN, W.; LI, Y. X.; ZHOU, Y. Y.; MA, Y. *et al.* Design, synthesis and SAR study of novel sulfonylurea derivatives containing arylpyrimidine moieties as potential anti-phytopathogenic fungal agents. **Chinese Chemical Letters**, 30, n. 12, p. 2160-2162, Dec 2019.

CHIAPINOTTO, D. M.; SCHAEGLER, C. E.; FERNANDES, J. P. S.; ANDRES, A. *et al.* CROSS-RESISTANCE OF RICE FLATSEED TO ALS-INHIBITING HERBICIDES. **Planta Daninha**, 35, 2017 2017.

CONSONNI, V.; BALLABIO, D.; TODESCHINI, R. Comments on the definition of the Q2 parameter for QSAR validation. **Journal of chemical information and modeling**, 49, n. 7, p. 1669-1678, 2009.

CORMANICH, R. A.; NUNES, C.; FREITAS, M. P. Chemical drawings correlate to biological properties: MIA-QSAR. **Química Nova**, 35, p. 1157-1163, 01/01 2012.

_____; FREITAS, M. P.; RITTNER, R. 2D Chemical Drawings Correlate to Bioactivities: MIA-QSAR Modelling of Antimalarial Activities of 2,5-Diaminobenzophenone Derivatives. **Journal of the Brazilian Chemical Society**, 22, n. 4, p. 637-642, 2011.

DAMALE, M. G.; HARKE, S. N.; KHAN, F. A. K.; SHINDE, D. B. *et al.* Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. **Mini reviews in medicinal chemistry**, 14, n. 1, p. 35-55, 2014.

DE OLIVEIRA JR, R. S.; CONSTANTIN, J.; INOUE, M. H. *Biologia e manejo de plantas daninhas. Curitiba, Brasil: Omnipax*, 2011.

DENNINGTON, R.; KEITH, T.; MILLAM, J. G.; VIEW, S. Gaussian, Inc. **Wallingford, CT**, 2008.

DU, X.; LI, Y.; XIA, Y.-L.; AI, S.-M. *et al.* Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. **International journal of molecular sciences**, 17, n. 2, p. 144-144, 2016.

DURIGON, M. R.; MARIANI, F.; SANTOS, F. M. D.; VARGAS, L. *et al.* Properties of the enzyme acetolactate synthase in herbicide resistant canola. **Bragantia**, 77, n. 3, p. 485-492, 2018-09 2018.

ERVIN, D. Herbicide Resistance: Economic and Environmental Challenges,” **AgBioForum** 19(2): 136-155. **AgBioForum**, 19, p. 136-155., 01/01 2016.

FERREIRA, M.; MONTANARI, C.; GAUDIO, A. Variable selection in QSAR. **Química Nova**, 25, p. 439-448, 05/01 2002.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. **Química nova**, 22, n. 5, p. 724-731, 1999.

FINNEY, D. J. **Probit analysis: a statistical treatment of the sigmoid response curve**. Cambridge university press, Cambridge, 1952.

FISER, A. Template-based protein structure modeling. **Computational biology**, p. 73-94, 2010.

FREITAS, M. P. MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzotrioles and their thio and sulfinyl congeners. **Organic & Biomolecular Chemistry**, 4, n. 6, p. 1154-1159, 2006.

_____ ; BROWN, S. D.; MARTINS, J. A. MIA-QSAR: a simple 2D image-based approach for quantitative structure–activity relationship analysis. **Journal of molecular structure**, 738, n. 1, p. 149-154, 2005.

GARCIA, M. D.; WANG, J. G.; LONHIENNE, T.; GUDDAT, L. W. Crystal structure of plant acetohydroxyacid synthase, the target for several commercial herbicides. **The FEBS journal**, 284, n. 13, p. 2037-2051, 2017.

GEE, S. K.; HAY, J. V. Recent Developments in the Chemistry of Sulfonylurea Herbicides. *In*: STETTER, J. (Ed.). **Herbicides Inhibiting Branched-Chain Amino Acid Biosynthesis: Recent Developments**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994. p. 15-46.

GOLBRAIKH, A.; TROPSHA, A. Beware of q². **Journal of molecular graphics & modelling**, 20, n. 4, p. 269-276, 2002.

GONÇALVES NETTO, A.; NICOLAI, M.; CARVALHO, S. J. P.; MALARDO, M. R. *et al.* Control of ALS- and EPSPS-Resistant *Amaranthus palmeri* by Alternative Herbicides Applied in PRE- and POST-Emergence. **Planta Daninha**, 37, 2019 2019.

GOODARZI, M.; FREITAS, M. P.; FERREIRA, E. B. Influence of Changes in 2-D Chemical Structure Drawings and Image Formats on the Prediction of Biological Properties Using MIA-QSAR. **QSAR & Combinatorial Science**, 28, n. 4, p. 458-464, 2009.

GUIDO, R. V.; ANDRICOPULO, A. D.; OLIVA, G. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estudos avançados**, 24, p. 81-98, 2010.

GUO, Y.; CHENG, L.; LONG, W. H.; GAO, J. Q. *et al.* Synergistic mutations of two rapeseed AHAS genes confer high resistance to sulfonylurea herbicides for weed control. **Theoretical and Applied Genetics**, 133, n. 10, p. 2811-2824, Oct 2020.

HENCKES, J. R.; CECHIN, J.; SCHMITZ, M. F.; PIASECKI, C. *et al.* Fitness Cost and Competitive Ability of Ryegrass Susceptible and with Multiple Resistance to Glyphosate, Iodosulfuron-Methyl, and Pyroxsulam. **Planta Daninha**, 37, 2019 2019.

HOOFT, R. W. W.; VRIEND, G.; SANDER, C.; ABOLA, E. E. Errors in protein structures. **Nature (London)**, 381, n. 6580, p. 272-272, 1996.

HOPFINGER, A. J.; WANG, S.; TOKARSKI, J. S.; JIN, B. *et al.* Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. **Journal of the American Chemical Society**, 119, n. 43, p. 10509-10524, 1997.

HUANG, M.; WANG, X.; DUAN, W. G.; LIN, G. S. *et al.* Synthesis, antifungal activity and 3D-QSAR study of novel acyl thiourea compounds containing gem-dimethylcyclopropane ring. **Molecular Diversity**, 2021. Article.

HUANG, X.; MILLER, W. A time-efficient, linear-space local similarity algorithm. **Advances in applied mathematics**, 12, n. 3, p. 337-357, 1991.

JANA, G. A.; DELGADO, E. J.; MEDINA, F. E. How Important Is the Synclinal Conformation of Sulfonylureas To Explain the Inhibition of AHAS: A Theoretical Study. **Journal of Chemical Information and Modeling**, 54, n. 3, p. 926-932, Mar 2014.

JARDIM, I. C. S. F.; ANDRADE, J. D. A.; QUEIROZ, S. C. D. N. D. Residues of pesticides in food: a global environmental preoccupation-Focussing on apples. **Química Nova**, 32, p. 996-1012, 2009.

KALSING, A.; GOULART, I. C. G. D. R.; MARIOT, C. H. P.; MENEZES, V. G. *et al.* Spatial and temporal evolution of imidazolinone-resistant red rice in 'Clearfield' rice cultivations. **Pesquisa Agropecuária Brasileira**, 54, 2019 2019.

KARAM, D.; DA SILVA, A. F.; GAZZIERO, D. L. P.; ADEGAS, F. S. *et al.* Situação atual da resistência de plantas daninhas a herbicidas nos sistemas agrícolas. **Embrapa Milho e Sorgo- Capítulo em livro científico (ALICE)**, 2018.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, 11, n. 1, p. 137-148, 1969.

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**, 20, n. 4, p. 770-787, 2009.

KRAEHMER, H.; VAN ALMSICK, A.; BEFFA, R.; DIETRICH, H. *et al.* Herbicides as Weed Control Agents: State of the Art: II. Recent Achievements. **Plant Physiology**, 166, n. 3, p. 1132-1148, 2014.

LAVORENTI, A.; PRATA, F.; REGITANO, J. B. Comportamento de pesticidas em solos: fundamentos. **Tópicos em ciência do solo**, 3, p. 430 il, 2003.

LI, J. H.; LI, R. H.; WANG, Y.; LI, S. X. *et al.* Synthesis, herbicidal activity, enzyme activity, and molecular docking of novel aniline thiourea. **Phosphorus, Sulfur and Silicon and the Related Elements**, 2021. Article.

_____; WANG, Y.; WU, Y. P.; LI, R. H. *et al.* Synthesis, herbicidal activity study and molecular docking of novel pyrimidine thiourea. **Pesticide Biochemistry and Physiology**, 172, 2021. Article.

LI, K. J.; QU, R. Y.; LIU, Y. C.; YANG, J. F. *et al.* Design, Synthesis, and Herbicidal Activity of Pyrimidine-Biphenyl Hybrids as Novel Acetohydroxyacid Synthase Inhibitors. **Journal of Agricultural and Food Chemistry**, 66, n. 15, p. 3773-3782, 2018. Article.

LIANG, Y.-F.; LONG, Z.-X.; ZHANG, Y.-J.; LUO, C.-Y. *et al.* The chemical mechanisms of the enzymes in the branched-chain amino acids biosynthetic pathway and their applications. **Biochimie**, 184, p. 72-87, 2021/05/01/ 2021.

LIU, P.; LONG, W. Current Mathematical Methods Used in QSAR/QSPR Studies. **International journal of molecular sciences**, 10, n. 5, p. 1978-1998, 2009.

MAO, B.; GAO, M.; CHEN, C.; LI, Z. *et al.* Design, synthesis and biological evaluation of novel N-nitrophenyl derivatives based on the structure of acetohydroxyacid synthase. **Pesticide Biochemistry and Physiology**, 145, p. 100-107, 2018. Article.

MARIANI, F.; SENSEMAM, S. A.; VARGAS, L.; AGOSTINETO, D. *et al.* Enzymatic Properties and Ryegrass Resistance Mechanism to Iodosulfuron-Methyl-Sodium Herbicide. **Planta Daninha**, 37, 2019 2019.

MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química nova**, 36, n. 4, p. 554-560, 2013.

MENDES, R. R.; ADEGAS, F. S.; TAKANO, H. K.; SILVA, V. F. V. *et al.* Multiple resistance to glyphosate and imazethapyr in *Bidens subalternans*. **Ciência e Agrotecnologia**, 43, 2019 2019.

MITRA, I.; SAHA, A.; ROY, K. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. **Molecular Simulation**, 36, n. 13, p. 1067-1079, 2010.

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

_____; FREITAS, M. P.; PINHEIRO, A. C. M.; BASTOS, S. C. Chemoface: a Novel Free User-Friendly Interface for Chemometrics. **Journal of the Brazilian Chemical Society**, 23, n. 11, p. 2003-2010, 2012.

PANDOLFO, C. E.; PRESOTTO, A.; POVERENE, M.; CANTAMUTTO, M. Limited occurrence of resistant radish (*Raphanus sativus*) to AHAS-inhibiting herbicides in Argentina. **Planta Daninha**, 31, n. 3, p. 657-666, 2013-09 2013.

POLANSKI, J. Receptor Dependent Multidimensional QSAR for Modeling Drug - Receptor Interactions. **Current medicinal chemistry**, 16, n. 25, p. 3243-3257, 2009.

QU, R. Y.; CAI, Z. M.; YANG, J. F.; LIU, Y. C. *et al.* Design, Synthesis and Biological Activity of Pyrimidyl-Salicylate Derivatives Containing Alkoxy Moiety. **Chinese Journal of Organic Chemistry**, 40, n. 11, p. 3953-3962, Nov 2020.

_____ ; YANG, J. F.; LIU, Y. C.; CHEN, Q. *et al.* Computational design of novel inhibitors to overcome weed resistance associated with acetohydroxyacid synthase (AHAS) P197L mutant. **Pest Management Science**, 73, n. 7, p. 1373-1381, 2017. Article.

RAMALHO, T. C.; FRANÇA, T. C. C.; RENNÓ, M. N.; GUIMARÃES, A. P. *et al.* Development of new acetylcholinesterase reactivators: Molecular modeling versus in vitro data. **Chemico-biological interactions**, 187, n. 1, p. 436-440, 2010.

RODRÍGUEZ TASSÉ, D.; BARBOSA GARCÍA, R. N.; RODRÍGUEZ VICENTE, E. Manejo de arvenses en caña de azúcar, impacto ambiental, efectividad económica y de control. **Centro Agrícola**, 46, n. 2, p. 64-71, 2019-06 2019.

ROY, K.; MITRA, I.; KAR, S.; OJHA, P. K. *et al.* Comparative Studies on Some Metrics for External Validation of QSPR Models. **Journal of chemical information and modeling**, 52, n. 2, p. 396-408, 2012.

RÜCKER, C.; RÜCKER, G.; MERINGER, M. γ -Randomization and Its Variants in QSPR/QSAR. **Journal of chemical information and modeling**, 47, n. 6, p. 2345-2357, 2007.

SANT'ANNA, C. M. R. Molecular modeling methods in the study and design of bioactive compounds: An introduction. **Revista Virtual de Química**, 1, n. 1, 2009.

SARMAH, A. K.; SABADIE, J. Hydrolysis of Sulfonylurea Herbicides in Soils and Aqueous Solutions: a Review. **Journal of agricultural and food chemistry**, 50, n. 22, p. 6253-6265, 2002.

TEÓFILO, R. Métodos Quimiométricos: Uma Visão Geral-Conceitos básicos de quimiometria. **Universidade Federal de Viçosa, Viçosa**, 1, p. 118, 2013.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. **Journal of medicinal chemistry**, 49, n. 11, p. 3315-3321, 2006.

TOBIAS, R. D. An Introduction to Partial Least Squares Regression. **Institute of Digital Research and Education**, 2016.

TODESCHINI, R.; CONSONNI, V.; MANNHOLD, R. Methods and principles in medicinal chemistry. **Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors. Wiley-VCH, Weinheim**, 2000.

TONG, J.; JIANG, G.; LI, L.; LI, Y. Molecular Virtual Screening Studies of Herbicidal Sulfonylurea Analogues Using Molecular Docking and Topomer CoMFA Research. **Journal of Structural Chemistry**, 60, n. 2, p. 210-218, 2019.

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

ULGUIM, A. R.; SILVA, B. M.; AGOSTINETTO, D.; AVILA NETO, R. C. *et al.* Resistance Mapping of the Genus *Cyperus* in Rio Grande do Sul and Selection Pressure Analysis. **Planta Daninha**, 37, 2019 2019.

VEDANI, A.; DOBLER, M. 5D-QSAR: The Key for Simulating Induced Fit? **Journal of medicinal chemistry**, 45, n. 11, p. 2139-2149, 2002.

_____; DOBLER, M.; LILL, M. A. Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor. **Journal of medicinal chemistry**, 48, n. 11, p. 3700-3703, 2005.

VON ZUBEN, M. O impacto econômico das pragas na lavoura. **AgroANALYSIS**, 39, n. 5, p. 45-45, 2019.

WANG, J.; YANG, Y.; HUANG, Y.; ZHANG, X. *et al.* Evaluation of modes of action of pesticides to *Daphnia magna* based on QSAR, excess toxicity and critical body residues. **Ecotoxicology and Environmental Safety**, 203, Oct 2020.

WATERHOUSE, A.; BERTONI, M.; BIENERT, S.; STUDER, G. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic acids research**, 46, n. W1, p. W296-W303, 2018.

WEI, W.; CHENG, D.; LIU, J.; LI, Y. *et al.* Design, synthesis and SAR study of novel sulfonylureas containing an alkenyl moiety. **Organic & Biomolecular Chemistry**, 14, n. 35, p. 8356-8366, 2016. 10.1039/C6OB01555G.

_____; ZHOU, S.; CHENG, D.; LI, Y. *et al.* Design, synthesis and herbicidal activity study of aryl 2, 6-disubstituted sulfonylureas as potent acetohydroxyacid synthase inhibitors. **Bioorganic & medicinal chemistry letters**, 27, n. 15, p. 3365-3369, 2017.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and intelligent laboratory systems**, 58, n. 2, p. 109-130, 2001.

WU, R.-J.; REN, T.; GAO, J.-Y.; WANG, L. *et al.* Chemical preparation, biological evaluation and 3D-QSAR of ethoxysulfuron derivatives as novel antifungal agents targeting acetohydroxyacid synthase. **European Journal of Medicinal Chemistry**, 162, p. 348-363, 2019/01/15/ 2019.

XAVIER, E.; TREZZI, M. M.; OLIVEIRA, M. C.; VIDAL, R. A. *et al.* Activity of Antioxidant Enzymes in *Euphorbia heterophylla* Biotypes and their Relation to Cross Resistance to ALS and Protox Inhibitors. **Planta Daninha**, 36, 2018 2018.

YOUSEFINEJAD, S.; HEMMATEENEJAD, B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. 149, p. 177-204, 2015.

ZHOU, S. A.; MENG, F. F.; HUA, X. W.; LI, Y. H. *et al.* Controllable Soil Degradation Rate of 5-Substituted Sulfonylurea Herbicides as Novel AHAS Inhibitors. **Journal of Agricultural and Food Chemistry**, 68, n. 10, p. 3017-3025, Mar 2020.

SEGUNDA PARTE

ARTIGO 1

Artigo aprovado e publicado na revista: Bulletin of Environmental Contamination and Toxicology

<https://doi.org/10.1007/s00128-022-03467-w>

QSAR-guided proposition of *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas as effective and safer herbicides

Abstract

Chlorinated agrochemicals play a major role in toxicity due especially to the labile C-Cl bond and high lipophilicity of organochlorines. In turn, urea and thiourea herbicides are widely used for weed control. A series of substituted *N*-benzoyl-*N'*-pyrimidin-2-yl thioureas has been recently synthesized and tested against *Brassica napus* L., demonstrating promising herbicidal activities, particularly for chlorinated derivatives. We have therefore modeled these activities using multivariate image analysis applied to quantitative structure-activity relationships (MIA-QSAR) to find out a significant and reliable correlation between measured and predicted inhibition of *B. napus* L. root growth (%) and, ultimately, to propose effective, non-chlorinated and/or less lipophilic *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thiourea candidates. The model was found to be predictive, giving an average r^2_{pred} in the external validation of 0.833. The predicted data for the proposed herbicides, interpreted in terms of MIA-plots of the chemical moieties responsible for bioactivity and supported by docking studies towards the photosystem II enzyme, suggest that substituents at both R¹ and R² positions modulate the agrochemical (R¹ = Cl increases and R² = OR decreases bioactivity) and environmental friendship (particularly with R² = OH) performances of this class of compounds.

Keywords: QSAR · docking · thioureas · herbicide · environment

Introduction

Ureas and thioureas are a group of herbicides used for general weed control in agricultural and nonagricultural practices (Gupta 2014). Examples of *N*-phenyl(thio)urea herbicides include the widely known diuron and methiuron compounds (Figure 1a), which are specific inhibitors of photosynthesis (by blocking the Q_B plastoquinone binding site of photosystem II) (Metz et al. 1986) and possible inhibitors of acetolactate synthase (an enzyme involved in the biosynthesis of branched amino acids) (Karasali and Maragou 2016), such as their sulfonylurea analogs. Some *N*-benzoyl-*N'*-pyrimidin-2-yl thioureas have been recently synthesized (Figure 1b) and a chlorinated derivative demonstrated to be the most promising candidate against *B. napus* L. (Li et al. 2021).

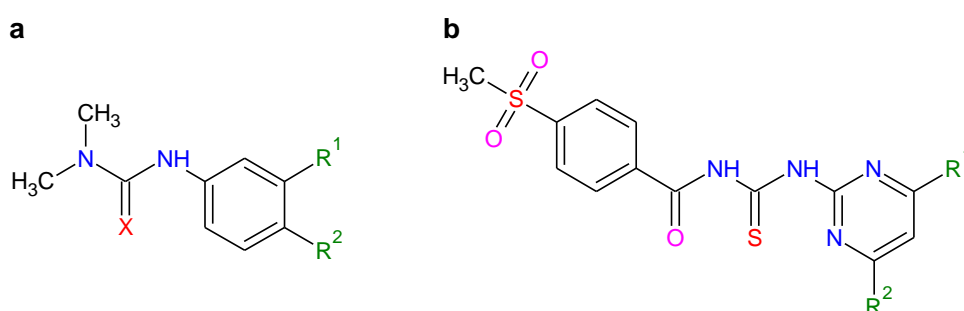


Figure 1. a) Commercial *N*-phenyl(thio)ureas, Diuron (X = O, R¹ and R² = Cl) and Methiuron (X = S, R¹ = CH₃ and R² = H), and b) *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas synthesized by Li et al. (2021), for which the derivative with R¹ = Cl and R² = OCH₃ demonstrated the best herbicidal performance.

Nevertheless, the reported risks of organochlorines to animal and human health appear to be associated with the relative ease for C-Cl cleavage (Liang et al. 2017). In addition, due to the large and diffuse nature of the chlorine atom, it is mostly responsible for the lipophilicity of organochlorine compounds, which is a major cause of biomagnification, bioconcentration, and bioaccumulation in fatty tissues of living organisms (Arnot and Gobas 2003; Moses et al. 2015; Thomann 1989). Therefore, the development of novel non-chlorinated and/or less lipophilic herbicide candidates would be valuable both to control resistant weeds and reduce the risks to the health and environment. Accordingly, this work reports a molecular modeling study to identify the chemical features of substituted *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas responsible for the herbicidal activity and, ultimately, proposing effective and safe candidates for use in agriculture.

First, a quantitative structure-activity relationship (QSAR) analysis based on multivariate image analysis (MIA) descriptors was applied to find a reliable correlation of the chemical structures of previously synthesized and tested *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas with the respective herbicidal activities. It is worth mentioning that QSAR has been widely accepted for *in silico* predictions to support of *in vitro* and *in vivo* studies (Barber et al. 2015). In turn, the MIA-QSAR method provides image contour maps that inform on relevant chemical moieties for the analyzed property (Barigye et al. 2016). Second, the mechanism of action of the known and proposed compounds was evaluated with the photosystem II (PSII) target through docking studies. Finally, a joint analysis of these results and the calculated lipophilicity described in terms of the octanol-water partition coefficient ($\log P$) allowed us to propose promising herbicide candidates.

Materials and Methods

The QSAR analysis employed 14 substituted *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas previously synthesized and tested against *B. napus* L., whose herbicidal activities are described in terms of average root growth inhibition (%) upon application of solutions at 1, 10, and 100 mg L⁻¹ concentrations (Li et al. 2021). The QSAR descriptors applied as predictors in the partial least squares (PLS) model are based on multivariate image analysis, corresponding to digital units (pixels) of 2D chemical images representing the molecular structures. These images were sketched using the GaussView program (Dennington et al. 2008) and are superimposable by the congeneric substructure corresponding to the 4-*N*-(methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thiourea moiety (Figure 2). The atoms in molecules were represented as circles with sizes proportional to the van der Waals radii to encode steric and hydrophobic properties. Also, considering that pixel colors correspond to RGB values, the color for each atom type was numbered proportionally to Pauling's electronegativity to encode dipolar and electrostatic interactions in the enzyme. Each image of dimension 200 × 228 pixels was unfolded to a row vector and then grouped to form a descriptor matrix of size 14 × (200 × 228) pixels. The calibration model was obtained by regression of the **X**-matrix against the **y**-block containing the herbicidal activity data. To assess whether the calibration result was obtained by chance correlation or not, a **y**-randomization test (10 runs) was performed and the unreal determination coefficient $r^2_{y\text{-rand}}$ was compared to the true r^2 using the c_r^2 parameter (values above 0.5 are acceptable) (Mitra et al. 2010). The absence of outliers was checked through William's plot of sample leverages × Student's residuals (Roy

et al. 2015). The model was both internally and externally validated using leave-one-out cross-validation and a bootstrapping procedure, respectively (Consonni et al. 2009; Kiralj and Ferreira 2009). The bootstrapping validation consisted of randomly selecting 10 test set compounds and then obtaining the statistical performance for each cycle. The QSAR development was entirely performed using the Chemoface program (Nunes et al. 2012). The lipophilicity parameter $\text{mlog } P$ was calculated using the Molinspiration program (www.molinspiration.com). The MolDock score 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 GEMDOCK program (Generic Evolutionary Method for molecular DOCK) with a new hydrogen bonding term and new charge schemes. MolDock SE was used as a search algorithm, and the maximum number of poses to generate was 100. The ligands and the amino acid side chains were flexible during the docking. Potential binding sites were identified using the built-in cavity detection algorithm. A grid covering the protein is created and at every grid point, a sphere is placed. It is checked whether this sphere will overlap with any of the spheres determined by the Van der Waals radii of the protein atoms. The cavities that are identified by the cavity detection algorithm are then used by the guided differential evolution search algorithm to focus the search, to that specific area during the docking simulation.

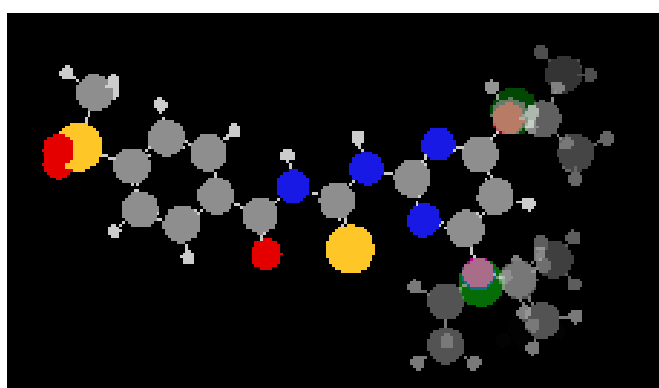


Figure 2. Superposition of images corresponding to the 14 4-*N*-(methanesulfonyl)benzoyl-(*N'*-pyrimidin-2-yl)thioureas used in the MIA-QSAR modeling. The ball-and-stick molecules were

sketched in the GaussView program (Dennington et al. 2008) and the atom sizes are proportional to the corresponding van der Waals radii. The atom colors (RGB pixel values) were renumbered to make correspondence with Pauling's electronegativity. Note that the congruent moieties are perfectly aligned, whereas the variable substituents at R¹ and R² positions explain the variance in the y-block.

A molecular docking study was conducted to explain the binding affinity of the 14 *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas against the PSII protein (EC:1.10.3.9). This structure exhibits a rich collection of atomic details: 16 chlorophylls, 4 pheophytins, 8 beta-carotenes, 2 plastoquinones, 2 protoporphyrin IX containing Fe, 2 Ca-Mn4-O5 clusters, 4 chloride ions, 2 Fe(II) ions, and 2 bicarbonate ions. The crystal structure of *B. napus* PSII (*Bn*PSII) was obtained by homology modeling (Fiser, 2010). The modeling started with a systematic search in the Swiss Bioinformatics Resource Portal (Expasy) to obtain the primary structure of the *Bn*PSII using the UniProt database (The UniProt Consortium, 2021) 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 (Huang and Miller, 1991). The templates were ranked based on the alignment score and the structural quality according to WHAT_CHECK (Hooft et al., 1996). The *Bn*PSII was constructed using the SWISS-MODEL server (Waterhouse et al., 2018). The prepared ligands were docked into the *Bn*PSII binding site using the Molegro Virtual Docker software (Thomsen and Christensen, 2006), which is designed for the prediction of the most likely binding conformation of a ligand towards a macromolecule (Ramalho et al., 2010).

Results and Discussion

The MIA-QSAR modeling was performed to the 14 *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas of Table 1. Once this series of compounds corresponds to a small data set, caution should be taken to interpret and validate the models. First, an applicability domain test was applied to evaluate the chemical space encompassing both the model descriptors and modeled response, and no outliers could be identified (Figure 3). Therefore, the MIA descriptors were regressed against the inhibition of *B. napus* L. root growth values (% , in decimal scale) using 2 PLS components, demonstrating a suitable modeling capability ($r^2 = 0.766$, RMSEC = 0.043). To check whether this correlation was not fortuitous, a *y*-

randomization test (10 cycles) was carried out and the resulting $r^2_{y\text{-rand}}$ of 0.396 was significantly lower than r^2 , as confirmed by the $^c r^2_p$ parameter of 0.533 ($^c r^2_p$ is a measure of how much r^2 differs from $r^2_{y\text{-rand}}$) (Mitra et al. 2010). To assess the predictive ability of the model, the leave-one-out cross-validation procedure was performed and a q^2 of 0.424 was obtained. However, a closer inspection of Figure 4 indicates that samples **4** and **13** possessing the highest and lowest (border) values of inhibition, respectively, influence considerably the regression line, as expected for a small data set; removal of these compounds from the plot increases the determination coefficient q^2 to 0.853.

Table 1. *N*-(4-methanesulfonyl)benzoyl-*N'*-(pyrimidin-2-yl)thioureas used in the MIA-QSAR modeling (and proposed analogs **P1-P3**), and the respective experimental inhibition of *B. napus* L. root growth (% , in decimal scale) and $\text{milog } P$ calculated using the Molinspiration program (www.molinspiration.com).

Compound	R ¹	R ²	Inhibition	$\text{milog } P$
1	OMe	OMe	0.5527	1.28
2	OEt	OEt	0.4558	2.04
3	O <i>i</i> Pr	O <i>i</i> Pr	0.4363	2.76
4	Cl	OMe	0.6913	1.91
5	Cl	OEt	0.4596	2.28
6	Cl	Cl	0.6240	2.53
7	Cl	NEt ₂	0.5907	2.70
8	OMe	NEt ₂	0.6001	2.08
9	OEt	NEt ₂	0.5224	2.46
10	O <i>i</i> Pr	NEt ₂	0.5109	2.82
11	Me	Cl	0.6359	1.95
12	Me	OMe	0.5227	1.32
13	Me	OEt	0.3384	1.70
14	OMe	OH	0.5660	1.01
P1	Me	NEt ₂	-	2.12
P2	Cl	OH	-	1.63
P3	Me	OH	-	1.05

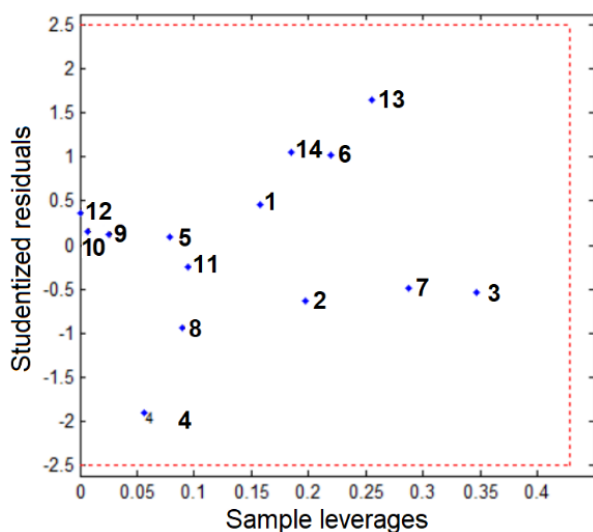


Figure 3. William's plot for outlier detection.

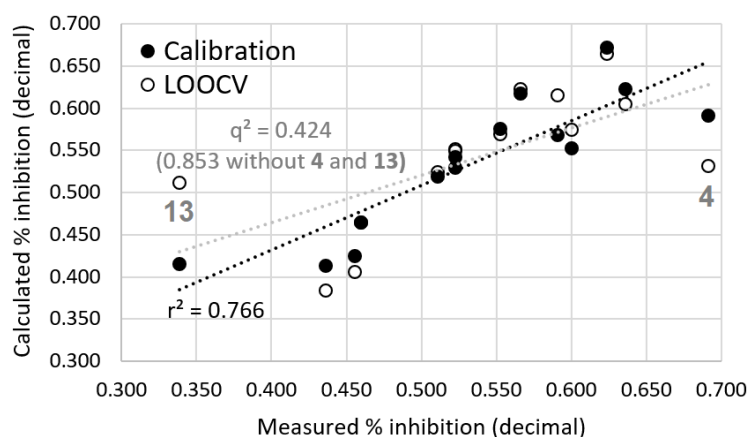


Figure 4. Correlation plot of measured vs. predicted inhibition of *B. napus* L. root growth (% , decimal scale) in calibration and leave-one-out cross-validation.

To completely validate the MIA-QSAR model, an external validation based on bootstrapping procedure was performed, in which four test set samples were randomly selected 10 times. This procedure is necessary for a small data set rather than a single splitting to avoid atypical behavior due to chance correlation and lack of robustness to resampling (Kiralj and Ferreira 2009). The results are given in Table 2, indicating that the MIA-QSAR model is reliable for prediction, once $r^2 > 0.6$ and $r^2_{\text{pred}} > 0.5$. In addition, the models demonstrated to be stable, because these metrics in the bootstrapping cycles oscillated little and above the acceptable cut-off values.

Table 2. Statistical performance of the MIA-QSAR modeling for external validation using the bootstrapping procedure (10 runs).^a

Parameter	1	2	3	4	5	6	7	8	9	10	Average
PLS components	3	3	2	2	2	3	1	4	2	4	2.6±1.0
RMSEC	0.015	0.037	0.049	0.049	0.046	0.031	0.057	0.012	0.050	0.026	0.037±0.016
r ²	0.975	0.870	0.717	0.754	0.777	0.892	0.654	0.983	0.735	0.930	0.829±0.116
RMSEP	0.141	0.074	0.063	0.041	0.049	0.097	0.035	0.109	0.054	0.081	0.074±0.034
r ² _{pred}	0.696	0.966	0.719	0.895	0.981	0.956	0.871	0.279	0.971	0.999	0.833±0.222
r ² _m	0.406	0.765	0.546	0.584	0.662	0.385	0.804	0.269	0.578	0.502	0.550±0.168

^a Root mean square errors in calibration (RMSEC) and prediction (RMSEP); r²_m is a measure of the proximity between the observed and predicted data, and values above 0.5 are considered acceptable (Roy et al., 2013).

The chemical features responsible for the pattern of % inhibition within the series of molecules can be investigated using the MIA-plots, which are based on contour heat maps of the PLS regression coefficients (**b**) and variable importance in projection (VIP) scores, and indicate how (increasing or decreasing) and how much each MIA descriptor influence the biological response (Figure 5). From the VIP plot, it is remarkable that the Cl substituent at R¹ and longer chains than OMe at R² mostly affect the % inhibition data. In turn, according to the **b**-plot, these groups contribute to increasing and decreasing the herbicidal activity, respectively. It is also worth noticing that OMe at R² is slightly disfavoring, whereas Cl at R² is slightly favoring for bioactivity. However, considering the balance between good herbicidal performance and low health/environmental risks, substituents possessing a neutral effect on the bioactivity and a less lipophilic character compared to Cl would be beneficial for novel herbicide candidates. Therefore, the unprecedented compounds **P1-P3** of Table 1 were proposed and their % inhibitions were estimated using the MIA-QSAR model. Accordingly, the calculated values for these herbicide candidates were 0.5188 (**P1**), 0.6333 (**P2**), and 0.5839 (**P3**). Whereas **P2** was found to be highly active and less lipophilic than **4** and **11** (the most active compounds within the series), the non-chlorinated **P3** presented low lipophilicity (due to the R² = OH) and higher calculated herbicidal potency than **14** (the least lipophilic compound). Overall, **P2** and **P3** seem to be promising thiourea-type herbicides.

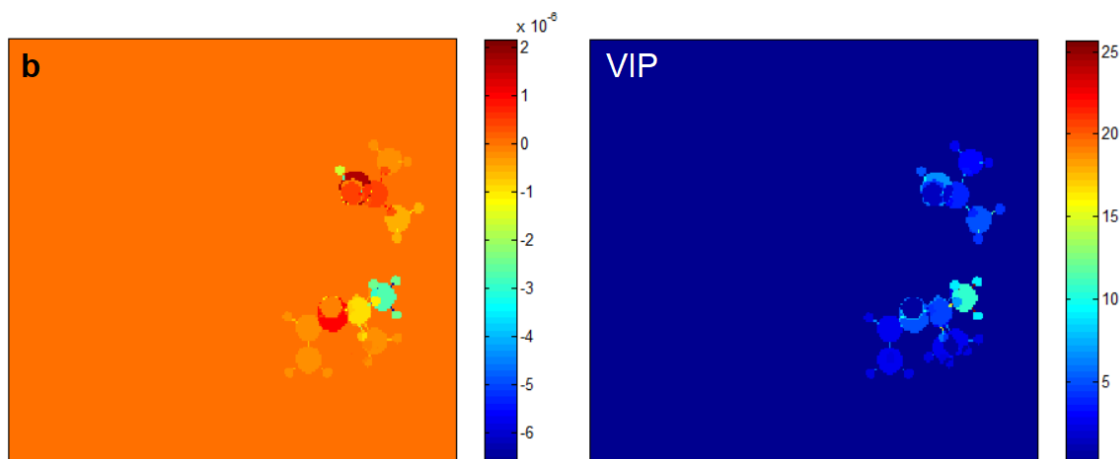


Figure 5. MIA-plots used to evaluate how (**b**-plot) and how much (VIP plot) the substituents in compounds **1-14** affect the % root growth inhibition.

The series of compounds **1-14** has been previously docked into the acetoxyacid synthase (AHAS) catalytic subunit binding cavity to study the mechanism of action of these compounds as herbicides (Li et al. 2021). The authors found that more hydrogen bonds between the ligand and the enzyme result in higher biological activity. However, the experimentally obtained AHAS inhibition values for these compounds do not properly correlate with the inhibition of *B. napus* L. root growth, once the linear regression between these two parameters yields an r^2 of *ca.* 0.46. Consequently, we suggest that photosystem II is the most likely target for ligand-enzyme interaction, such as for the usual atrazine and phenylurea herbicides (Liu 2010), and not AHAS, which is the main enzyme target for sulfonylureas (Brown and Cotterman 1994). Accordingly, further efforts to understand the mechanism of action of **1-14**, as well as of the proposed compounds **P1-P3**, were carried out over the photosystem II enzyme.

To obtain the crystal structure of *BnPSII*, the primary sequences of *B. napus* and *Cyanidium caldarium* were aligned and demonstrated 89.5% identity. Therefore, considering that aligned sequences over 40-50% are nearly identical, the alignments were then considered accurate in a structural sense (Orry and Abagyan, 2012). The SWISS-MODEL server was used to make the protein modeling and the PDB-entry 4YUU was the template selected through the sequence alignment (Ago et al., 2016). The quality of the homology model was evaluated by different tools, such as the Ramachandran plot (Figure 6). The modeled protein represents 98.6% of overall residues in the most favored regions, and 1.4% in allowed regions, representing a good quality model. The RMSD value of 1.46 Å was found in the best superimposition for protein structures; lower values than 2 Å mean better alignment between the structures.

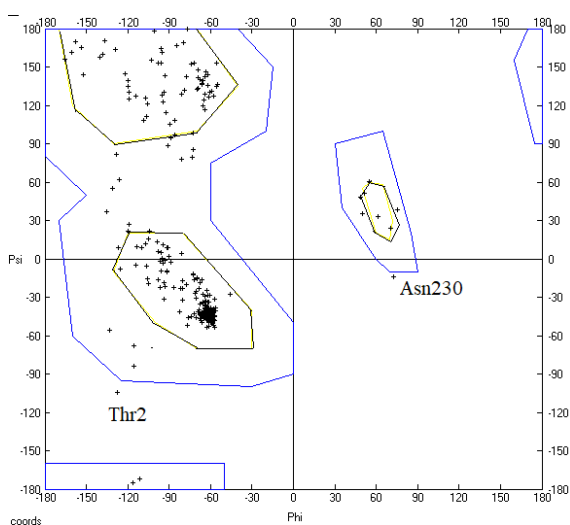


Figure 6. Ramachandran plot showing the homology results.

To better understand the interaction mode of this class of compounds, the molecules were docked inside the binding pocket of the *Bn*PSII enzyme obtained by homology modeling. It has been confirmed that PSII is the most likely target instead of AHAS, once the docking scores for the 14 compounds correlate in a high extent to the inhibition growth values ($r^2 = 0.73$). A subset region around the active site was defined by the following amino acid residues: Phe211, Met214, His215, Leu218, Ile248, Ala251, His252, Phe255, Ser264, Ser264, Leu271, Phe274, and 275. The poses were determined as the highest energy scored during docking. The energy values are given in Table 3. Amongst the compounds of Table 1, ligands **4** and **11** formed more stable complexes with the *Bn*PSII enzyme. The proposed compounds **P1**, **P2**, and **P3** exhibited score energy values similar to that of compounds **4** and **11**. Hydrogen bonds are a primary contributor to the binding affinity of ligands with receptor; in our case, all compounds form hydrogen bonds with Ser264. In addition, π -sigma or π -alkyl interactions have been observed between the 4-(methanesulfonyl)benzoyl group and Met214, Leu218, and Leu271, in addition to π - π stacking interactions with Phe255. The pyrimidinyl ring interacts by π - π stacking with Phe274. The optimum structures of compounds **4** and **P1** (the best herbicide candidate according to the docking studies), and the main amino acid residues in the *Bn*PSII binding pocket are illustrated in Figure 7.

Table 3. Energy values (kcal mol⁻¹) obtained from the docking between *Bn*PSII enzyme and the studied compounds.

Compounds	MolDock Score	Intermolecular energy	Hydrogen bond energy
1	-99.83	-122.15	-2.72
2	-73.30	-107.12	-3.97
3	-87.67	-124.56	-3.77
4	-108.52	-132.31	-3.74
5	-67.74	-97.40	-3.95
6	-94.17	-112.10	-2.21
7	-101.14	-122.35	-2.44
8	-103.07	-130.70	-2.69
9	-89.70	-123.38	-4.01
10	-79.01	-109.49	-3.73
11	-105.22	-126.13	-2.90
12	-96.20	-115.95	-3.65
13	-69.42	-98.93	-1.94
14	-89.00	-109.91	-3.75
P1	-107.62	-131.25	-2.90
P2	-103.29	-126.03	-3.35
P3	-103.76	-125.69	-3.76

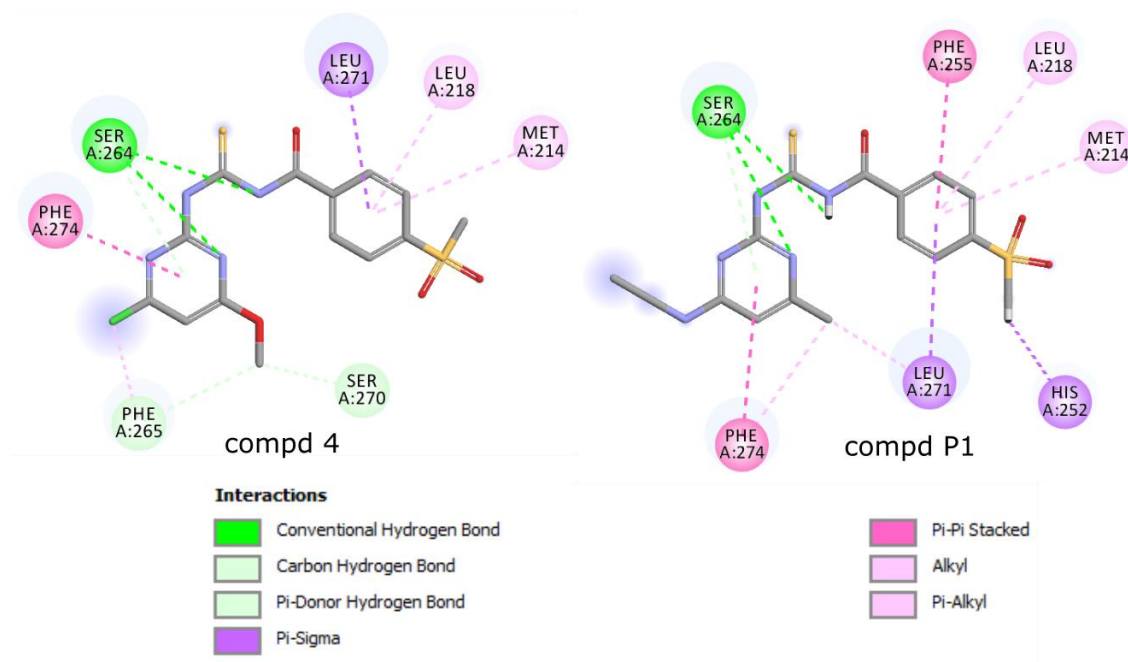


Figure 7. Important ligand-enzyme interactions obtained from the docking studies for compounds **4** and **P1**.

In summary, at least three candidates demonstrated promising herbicide-likeness, including with respect to environmental issues. Nevertheless, synthesis and biological assays are further required to confirm our computational outcomes.

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

Ago H, Adachi H, Umena Y, Tashiro T, Kawakami K, Kamiya N, Tian L, Han G, Kuang T, Liu Z, Wang F, Zou H, Enami I, Miyano M, Shen J-R (2016) Novel features of eukaryotic photosystem II revealed by its crystal structure analysis from a red alga. *J Biol Chem* **291**:5676-5687. <https://doi.org/10.1074/jbc.M115.711689>

- Arnot JA, Gobas FAPC (2003) A generic QSAR for assessing the bioaccumulation potential of organic chemicals in aquatic food webs. *QSAR Comb Sci* 22:337-345. <https://doi.org/10.1002/qsar.200390023>
- Barber C, Amberg A, Custer L, Dobo KL, Glowienke S, Gompel JV, Gutsell S, Harvey J, Honma M, Kenyon MO, Kruhlak N, Muster W, Stavitskaya L, Tealsdale A, Vessey J, Wichard J (2015) Establishing best practise in the application of expert review of mutagenicity under ICH M7. *Reg Toxicol Pharmacol* 73:367-377. <https://doi.org/10.1016/j.yrtph.2015.07.018>
- Barigye SJ, Duarte MH, Nunes CA, Freitas MP (2016) MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. *RSC Adv* 6:49604-49612. <https://doi.org/10.1039/C6RA09593C>
- Brown HM, Cotterman JC (1994) Recent advances in sulfonylurea herbicides. In: Stetter J (ed) *Herbicides inhibiting branched-chain amino acid biosynthesis. Chemistry of plant protection*, vol 10. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-79107-9_4
- Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q^2 parameter for QSAR validation. *J Chem Inf Model* 49:1669-1678. <https://doi.org/10.1021/ci900115y>
- Dennington II RD, Keith TA, Millam JM (2008) *GaussView*, Gaussian, Inc, Wallingford
- Fiser A (2010) Template-based protein structure modeling. In: Fenyő D (ed) *Computational Biology. Methods in Molecular Biology (Methods and Protocols)*. Humana Press, Totowa, pp 73-94.
- Gupta PK (2014) Herbicides and fungicides. In: Gupta RC (ed) *Biomarkers in toxicology*. Academic Press, Cambridge, pp 409-431
- Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272. <https://doi.org/10.1038/381272a0>
- Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. *Adv Appl Math* 12:337-357. [https://doi.org/10.1016/0196-8858\(91\)90017-D](https://doi.org/10.1016/0196-8858(91)90017-D)
- Karasali H, Maragou N (2016) Pesticides and herbicides: Types of pesticide. In: Caballero B, Finglas PM, Toldrá F (eds) *Encyclopedia of food and health*. Elsevier, Amsterdam, pp 319-325

- Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *J Braz Chem Soc* 20:770-787. <https://doi.org/10.1590/S0103-50532009000400021>
- Li J-H, Wang Y, Wu Y-P, Li R-H, Liang S, Zhang J, Zhu Y-G, Xie B-J (2021) Synthesis, herbicidal activity study and molecular docking of novel pyrimidine thiourea. *Pest Biochem Physiol* 172:104766. <https://doi.org/10.1016/j.pestbp.2020.104766>
- Liang X, Huang T, Li M, Mack J, Wildervanck M, Nyokong T, Zhu W (2017) Highly efficient C–Cl bond cleavage and unprecedented C–C bond cleavage of environmentally toxic DDT through molecular electrochemical catalysis. *Appl Cat A* 545:44-53. <https://doi.org/10.1016/j.apcata.2017.07.026>
- Liu J (2010) Phenylurea herbicides. In: Krieger R (ed) *Hayes' Handbook of Pesticide Toxicology*. Academic Press, Cambridge, pp 1725-1731
- Metz J, Pakrasi H, Seibert M, Arntzer C (1986) Evidence for a dual function of the herbicide-binding D1 protein in photosystem II. *FEBS Lett* 205:269-274. [https://doi.org/10.1016/0014-5793\(86\)80911-5](https://doi.org/10.1016/0014-5793(86)80911-5)
- Mitra I, Saha A, Roy K (2010) Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol Simul* 36:1067-1079. <https://doi.org/10.1080/08927022.2010.503326>
- Moses SK, Harley JR, Lieske CL, Muir DCG, Whiting AV, O'Hara TM (2015) Variation in bioaccumulation of persistent organic pollutants based on octanol–air partitioning: Influence of respiratory elimination in marine species. *Mar Poll Bull* 100:122-127. <https://doi.org/10.1016/j.marpolbul.2015.09.020>
- Nunes CA, Freitas MP, Pinheiro ACM, Bastos SC (2012) Chemoface: A novel free user-friendly interface for chemometrics. *J Braz Chem Soc* 23:2003–2010. <https://doi.org/10.1590/S0103-50532012005000073>
- Orry AJW, Abagyan R (2012) *Homology Modeling: Methods and Protocols*. Humana Press, Totowa. <https://doi.org/10.1007/978-1-61779-588-6>
- Ramalho TC, França TCC, Rennó MN, Guimarães AP, da Cunha EFF, Kuca K (2010) Development of new acetylcholinesterase reactivators: molecular modeling versus in vitro data. *Chem Biol Interact* 185:73-77. <https://doi.org/10.1016/j.cbi.2010.05.016>

- Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) 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* 34:1071-1082. <https://doi.org/10.1002/jcc.23231>
- Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Sys* 145:22-29. <https://doi.org/10.1016/j.chemolab.2015.04.013>
- The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480-D489. <https://doi.org/10.1093/nar/gkaa1100>
- Thomann RV (1989) Bioaccumulation model of organic chemical distribution in aquatic food chains. *Environ Sci Technol* 23:699-707. <https://doi.org/10.1021/es00064a008>
- Thomsen R, Christensen MH (2006) MolDock: A new technique for high-accuracy molecular docking. *J Med Chem* 49:3315-3321. <https://doi.org/10.1021/jm051197e>
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296-W303. <https://doi.org/10.1093/nar/gky427>

ARTIGO 2

Artigo aprovado e publicado na revista: New Journal of Chemistry

DOI: 10.1039/d2nj02087d

Quantitative structure-activity relationship modeling of the inhibitory activities of sulfonylurea herbicides and proposition of new derivatives

Natânia E. Rodrigues,^a Matheus P. Freitas^{a,*}

^a Department of Chemistry, Institute of Natural Sciences, Federal University of Lavras, 37200-900, Lavras, MG, Brazil.

Abstract

The quantitative structure-activity relationship study of the herbicidal activity of a series of sulfonylurea derivatives against *Brassica napus* is reported herein. A few multivariate image analysis (MIA) descriptors have been selected through genetic algorithm from a pool of pixels renumbered to encode atomic properties, namely the van der Waals radii and electronegativity. These descriptors correlate with the *B. napus* growth inhibition to a great extent ($r^2 = 0.90 \square 0.91$, $q^2 = 0.76 \square 0.82$) and, therefore, could be used to estimate the pIC₅₀ values of hitherto unknown sulfonylurea analogs (IC₅₀ = half maximal inhibitory concentration). The design of these novel derivatives was guided by MIA contour maps that reveal the substituent effects on the response variable. As a result, four agrochemical candidates were proposed and three of them exhibited calculated pIC₅₀ values higher than that of chlorsulfuron and the data set compounds. This may be an important step to drive the synthesis of new sulfonylurea herbicides with improved properties against resistant weeds.

1. Introduction

Sulfonylurea herbicides have been increasingly marketed in comparison to other classes of compounds, *e.g.* dinitroanilines, ureas, glyphosate, and triazines, due to their advantages in terms of efficacy and environmental friendship.¹ For example, degradation of sulfonylureas in soil is pH dependent and, accordingly, persistence of these compounds in acidic soils is relatively low due to rapid chemical hydrolysis, while in neutral- to low-alkaline soils chemical hydrolysis is slow.² These compounds act by inhibiting the acetohydroxyacid synthase (AHAS) enzyme, exhibiting successful performance against both monocotyledonous and dicotyledonous weeds.³ Chlorsulfuron (Figure 1) is usually the control compound for development studies, but resistance against well-known herbicides, such as chlorsulfuron, is a major problem faced by agrochemical candidates.⁴ Therefore, the design of new herbicides is an essential task before synthesizing and testing molecule candidates, thus avoiding trial and error attempts.

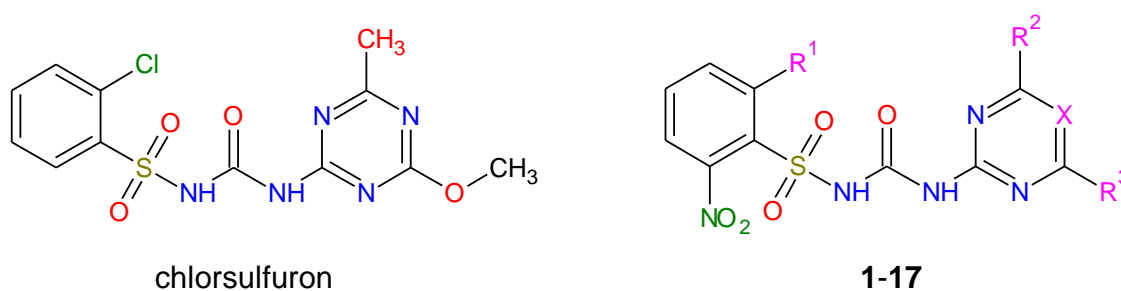


Figure 1. Chlorsulfuron and other sulfonylureas studied herein, as synthesized by Wei et al.² X, R¹, R², and R³ substituents are given in Table 1.

A series of 17 sulfonylureas has been recently synthesized and assayed using *Brassica napus* (Figure 1), giving inhibition rates varying from 0% at 0.1 mg L⁻¹ to 72.6% at 10 mg L⁻¹, while chlorsulfuron presented 68.9% of inhibition at the highest concentration.³ After the conversion of the inhibition rates to pIC₅₀ using the probit method,⁵ the most promising derivatives were found to be those with Cl or OCH₂CF₃ substituents at the R¹ position, and CH₃ or OCH₃ substituents at the R² and R³ positions (Table 1). The probit method analyzes the relationship between a stimulus (dose) and the quantal response, assuming that the percent response is related to the log dose as the cumulative normal distribution; a few doses are required to obtain a reliable result and, in the present case, the inhibitory rates against *B. napus* (%) obtained for three concentrations (0.1, 1, and 10 mg L⁻¹) were converted to the half maximal inhibitory concentration (IC₅₀, in mol L⁻¹) and, then, to -log IC₅₀ (pIC₅₀). The logarithmic basis

allows for an optimal range for the biological response, which spanned from 4.25 to 8.92 in the present case.⁶

To obtain the sulfonylurea derivatives, Wei et al.³ used a many-steps synthetic route starting from a 1,2,3-tri-substituted benzene (substituents NO₂, SBn, and R¹ = halogen), followed by conversion of the SBn group into pyrimidinyl or triazinyl sulfonylurea moieties. Conventional herbicidal sulfonylureas require an electron-withdrawing group at the sulfonyl-*ortho* position⁷ and, accordingly, Wei et al.³ have found that a F atom was the most favorable substituent at 2-position of the aryl moiety, but other substituents did not show any obvious priority. Therefore, a systematic evaluation using quantitative modeling approaches may aid a better understanding of the substituent effects ruling the biological activity.

Quantitative structure-activity relationships (QSAR) modeling may be properly used to confirm and interpret these findings, as well as to predict the pIC₅₀ of novel proposed derivatives. QSAR is the final result of the process that starts with a suitable description of molecular structures and ends with some inference, hypothesis, and prediction on the behavior of molecules in environmental, biological, and physicochemical systems in analysis.⁸ This work reports the QSAR modeling of the herbicidal activities of the series of sulfonylureas mentioned above using multivariate image analysis (MIA) molecular descriptors.⁹ Tentatively, the MIA-QSAR models obtained were applied to estimate the pIC₅₀ values of rationally designed sulfonylurea candidates.

Table 1. The series of sulfonylureas and respective pIC₅₀ values (IC₅₀ in mol L⁻¹).^a

Compound	X	R ¹	R ²	R ³	pIC ₅₀
10a	CH	Cl	CH ₃	CH ₃	6.35
10b	CH	Cl	OCH ₃	OCH ₃	8.92
10c	CH	Cl	H	CH ₃	5.20
10d	CH	OCH ₂ CF ₃	CH ₃	CH ₃	8.65
10e	CH	Cl	CH ₃	OCH ₃	5.41
10f	CH	Cl	Cl	OCH ₃	5.83
10g	N	Cl	CH ₃	OCH ₃	5.13
10h	CH	F	OCH ₃	OCH ₃	6.26
10i	CH	F	CH ₃	OCH ₃	6.54
10j	CH	F	H	CH ₃	5.27
10k	N	F	CH ₃	OCH ₃	4.25
10l	CH	I	CH ₃	OCH ₃	6.35
10m	CH	I	OCH ₃	OCH ₃	5.57
10n	CH	I	H	CH ₃	5.11
10o	CH	Br	CH ₃	OCH ₃	6.40
10p	CH	Br	OCH ₃	OCH ₃	5.46
10q	N	Br	CH ₃	OCH ₃	5.22

^a pIC₅₀ of chlorsulfuron: 6.65 (based on the conversion of inhibition rates available in Wei *et al.*³).

Material and Methods

The series of 17 sulfonylureas (**10a**–**10q**) of Figure 1 and Table 1 was retrieved from the literature,³ whose inhibition rates were obtained in three concentrations (0.1, 1.0, and 10 mg L⁻¹) and expressed in terms of percentage of *B. napus* growth. These concentrations were

converted to mol L^{-1} and then to pIC_{50} using the probit method.⁵ These data were regressed against selected MIA descriptors using multiple linear regression. The MIA descriptors were generated from molecular images sketched using the GaussView program,¹⁰ where atoms are represented as colored circles with sizes proportional to the respective van der Waals radii. Since colors are described by the RGB system, whose numeric values range from zero (black) to 765 (white), the pixels for each atom type were renumbered to make a correspondence with the respective van der Waals radius (r_{vdw}) and electronegativity (ϵ) values. Each image of dimension 300×354 pixels was unfolded to a row vector of 106,200 pixels and grouped to obtain a $17 \times 106,200$ data matrix. The most relevant MIA descriptors were selected through genetic algorithm using a fitness function based on error/MAE-based parameters and 100 iterations/generations, mutation probability of 0.3, and 30 equations selected in each generation.¹¹ As a result, five selected descriptors for each type of MIA descriptors ($\text{MIA}_{r_{\text{vdw}}}$ and MIA_{ϵ}) were regressed against the pIC_{50} values using multiple linear regression (MLR) to obtain the MIA-QSAR models (Figure 2). To attest to the predictive performance of these models, the \mathbf{X} matrices of 17×5 dimension were split into training (13 samples) and test (4 samples) sets using Kennard-Stone¹² and random sampling. The resulting models were validated according to the best practices in QSAR.¹³ The original descriptor matrices were also regressed against the pIC_{50} values using partial least squares (PLS) regression¹⁴ to generate variable importance in projection (VIP) scores and PLS regression coefficients (\mathbf{b}), which were applied to build MIA-plots¹⁵ to understand how the substituents affect the response variables. The VIP scores describe the effect of each variable on the PLS model, and are computed as the weighted sum of squares of the weights, which measures the percentage of the \mathbf{y} variance explained with each latent variable.¹⁶ In turn, \mathbf{b} corresponds to the vector of PLS inner relation coefficients and gives the importance of each original variable to \mathbf{y} .¹⁷ This procedure was performed using the Chemoface program.¹⁸

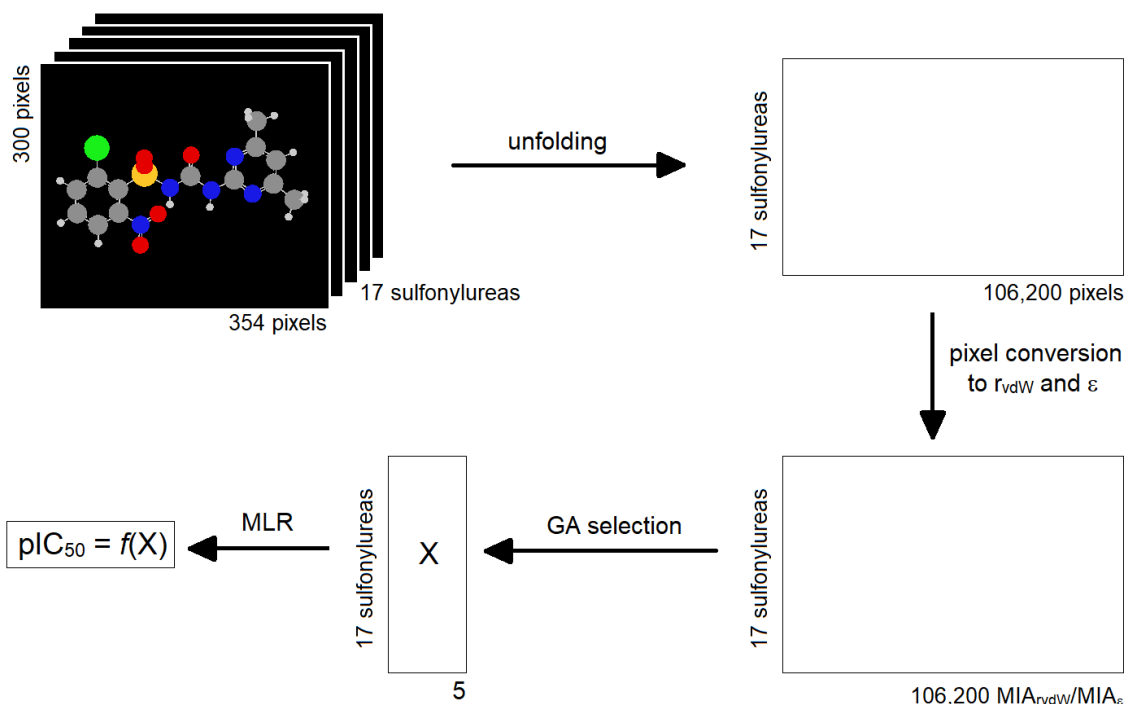


Figure 2. Workflow followed to obtain the MIA-QSAR models.

Results and Discussion

Careful validation of the QSAR models should be performed as the data set is composed of a small amount of only 17 compounds, with an activity spread of three to four orders of magnitude. This is noteworthy and unveils the strong effect of particular groups attached to the common scaffold. First, to attest whether these compounds are all encompassed by the same sampling space, an applicability domain test using studentized residuals and sample leverages was performed, and any outlier was found (Figure 3). Subsequently, the data matrices composed of 5 selected MIA_{r_{vdw}} and MIA_ε descriptors each (Table 2 and Figure 4) were regressed against the inhibition data (pIC_{50}), and both descriptor types demonstrated good modeling performance ($r^2 > 0.90$), as well as high determination coefficients in the leave-one-out cross-validation ($q^2 > 0.76$) (Figure 5).¹⁹ These good correlations were not fortuitous, as demonstrated by the ${}^c r_p^2 = 0.74$ obtained in the y-randomization tests (mean of 10 repetitions).²⁰ This demonstrates that the MIA descriptors encode the same steric and electrostatic properties of molecular field-based methods, such as the well-established CoMFA.²¹ The selected descriptors lie in the region of X, R¹, R², and R³ substituents, but especially R² (Figure 4). Particular emphasis should be given to MIA_ε1739, which reflects the importance of a fluorinated group at the R¹ position in

compound **10d**, and the effect of $MIA_{\epsilon}4597$ that differentiates pyrimidinyl and triazinyl compounds. The obtained models are described in Eqs. (1) and (2).

$$pIC_{50} = 3.67 - 0.0015 \times MIA_{rdw}3608 - 0.0031 \times MIA_{rdw}4858 + 0.0090 \times MIA_{rdw}5271 - 0.0189 \times MIA_{rdw}5869 + 0.0253 \times MIA_{rdw}6047 \quad (1)$$

$N = 17$, $r^2 = 0.909$, $RMSEC = 0.356$, $q^2 = 0.824$, $RMSECV = 0.514$, $r^2_{y\text{-rand}} = 0.312$, $RMSEC_{y\text{-rand}} = 0.971$, $^c r^2_p = 0.736$.

$$pIC_{50} = 12.28 + 0.0062 \times MIA_{\epsilon}1739 + 0.0005 \times MIA_{\epsilon}3374 - 0.0150 \times MIA_{\epsilon}3384 + 0.0157 \times MIA_{\epsilon}4075 - 0.0267 \times MIA_{\epsilon}4597 \quad (2)$$

$N = 17$, $r^2 = 0.904$, $RMSEC = 0.364$, $q^2 = 0.763$, $RMSECV = 0.506$, $r^2_{y\text{-rand}} = 0.297$, $RMSEC_{y\text{-rand}} = 0.984$, $^c r^2_p = 0.741$.

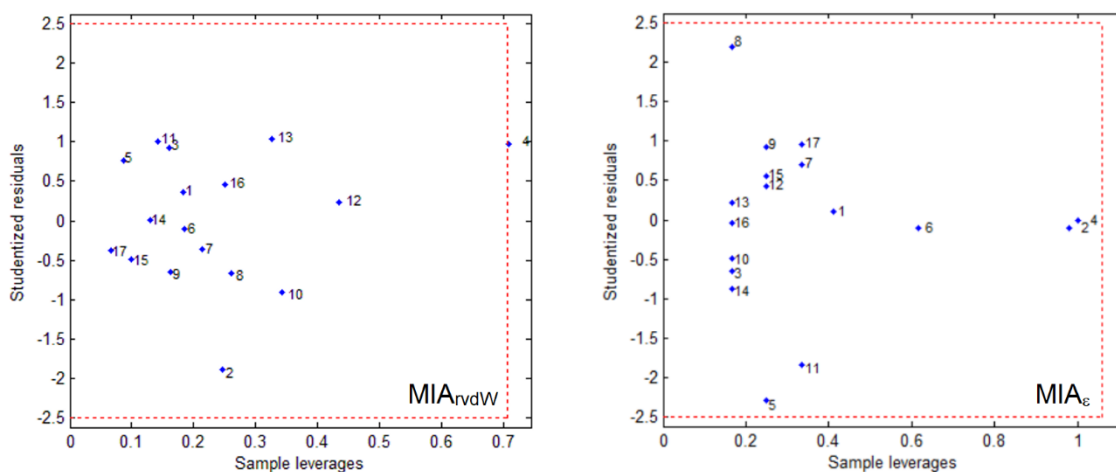


Figure 3. William's plot of studentized residuals \times sample leverages for applicability domain test.

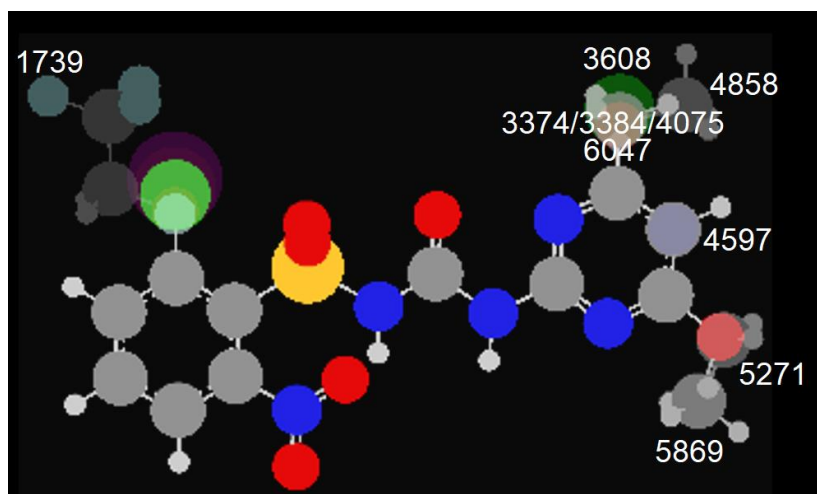


Figure 4. Approximate coordinates of the selected MIA descriptors.

Table 2. Descriptor matrices used in the MIA-QSAR modeling using multiple linear regression. The descriptor numbering refers to pixel coordinates after unfolding the images to row vectors.

Cpd	MIA _{rvdw}					MIA _ε				
	3608	4858	5271	5869	6047	1739	3374	3384	4075	4597
10a	0	0	0	0	120	0	255	220	255	255
10b	0	170	170	0	170	0	0	0	220	255
10c	0	0	170	0	0	0	0	0	0	255
10d	0	0	170	0	120	398	0	220	255	255
10e	0	0	170	170	170	0	0	220	255	255
10f	289	0	170	170	170	0	316	316	316	255
10g	0	0	0	170	170	0	0	220	255	304
10h	0	170	170	170	170	0	0	0	0	255
10i	0	0	170	170	170	0	0	220	255	255
10j	0	0	170	0	0	0	0	0	0	255
10k	0	0	0	170	170	0	0	220	255	304
10l	0	0	170	170	170	0	0	220	255	255
10m	0	170	170	170	170	0	0	0	0	255
10n	0	0	170	0	0	0	0	0	0	255
10o	0	0	170	170	170	0	0	220	255	255
10p	0	170	170	170	170	0	0	0	0	255
10q	0	0	0	170	170	0	0	220	255	304

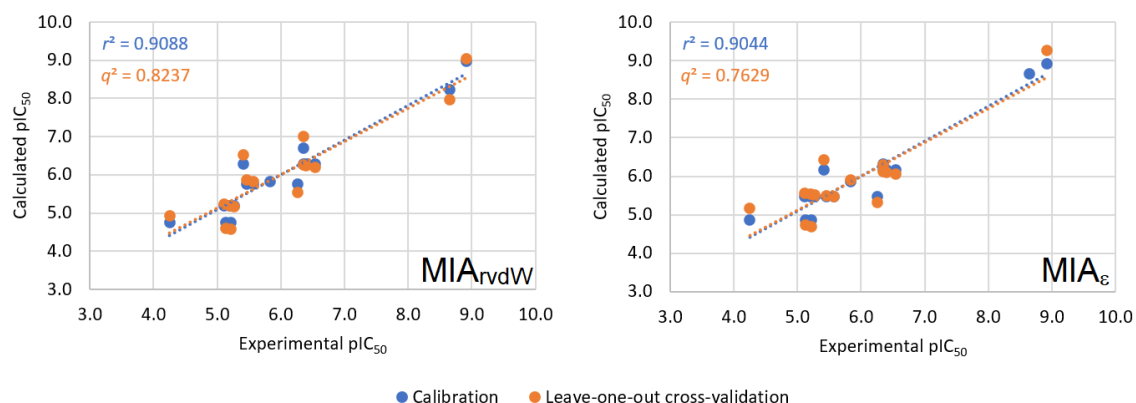


Figure 5. Scattering plots of experimental \times calculated pIC₅₀ obtained in the MIA-QSAR modeling.

To completely assure the predictive capability of the models, the data set was divided into training and test subsets for external validation purposes, using two approaches: Kennard-Stone¹² and random sampling. The former selected the same calibration and test sets for both models, while the latter is expected to give different sets after two runs. The resulting models were found to be highly predictive (r^2_{pred} and $r^2_{\text{m}} \gg 0.5$),²² especially for the randomly selected samples (Table 3). Considering that the data could be reliably modeled, the substituent effects were interpreted using MIA plots, namely the maps of variable importance in projection (VIP) scores and PLS regression coefficients (**b**) (Figure 6). The former informs on the substituents most affecting the response variable (from nothing – dark blue – to much – dark red). In turn, the latter indicates if the substituents contribute to increasing (tending to dark red) or decreasing (tending to dark blue) the pIC₅₀ values.

Table 3. Statistical results obtained for the MIA-QSAR models using training/test set splitting based on Kennard-Stone (K-S) and random sampling.

Parameter	MIA _{r_{vdW}}		MIA _{ϵ}	
	K-S (10n,10o,10p,10q)	Random (10c,10d,10k,10l)	K-S (10n,10o,10p,10q)	Random (10i,10m,10o,10q)
RMSEC	0.362	0.331	0.374	0.361
r^2	0.921	0.888	0.915	0.924
RMSE _{y-} rand	1.003	0.700	1.021	0.978
$r^2_{y\text{-rand}}$	0.374	0.475	0.348	0.431
$^c r^2_p$	0.709	0.606	0.720	0.675
RMSECV	0.597	0.549	0.370	0.756
q^2	0.809	0.973	0.511	0.863
RMSEP	0.391	0.558	0.611	0.623
r^2_{pred}	0.617	0.973	0.717	0.709
r^2_m	0.603	0.560	0.388	0.500

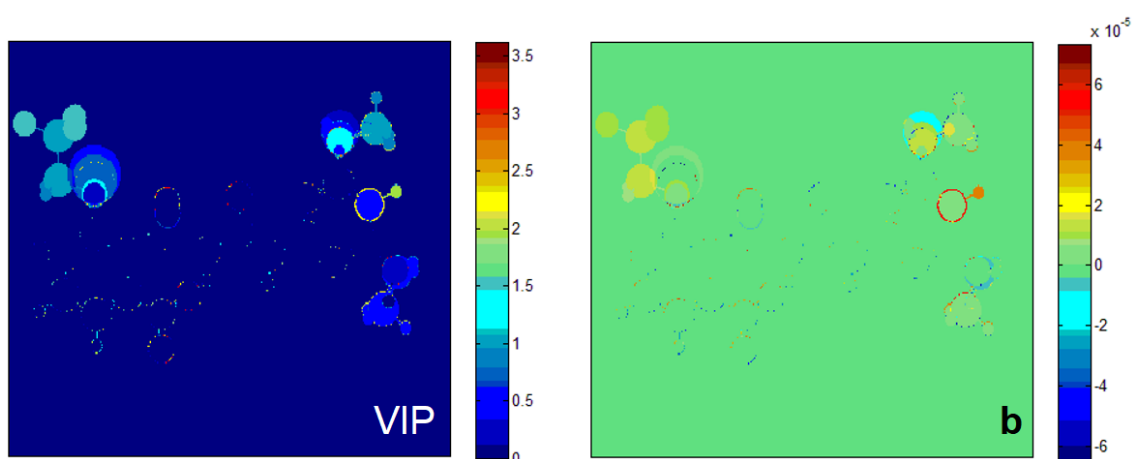


Figure 6. MIA plots used to interpret the substituent effects on pIC₅₀.

The VIP plot indicates that regions most affecting the pIC_{50} data lie on X, but also R^1 and R^2 , since the substituent effect ranges from dark blue (unimportant) to dark red (very important). According to the **b** plot, dark blue regions indicate negative influence on pIC_{50} , while the opposite effect is observed in dark red regions. Therefore, X = CH is more beneficial than N, while OCH_2CF_3 in R^1 rather than halogens favor the biological activity. Substituents at R^2 influence the target binding at atomic scale in a smaller extent than X and R^1 . Consequently, it follows that pyrimidine are more promising than triazine derivatives, while large and electron withdrawing groups in the R^1 position are more favorable than halogens. This agrees with similar aniline and pyrimidine thioureas previously synthesized by Li *et al.*,^{23,24} according to the authors, electronegative groups attached to the aromatic rings, such as OCH_3 , CF_3 , and SO_2CH_3 , participate in hydrogen bonds that stabilize the ligand-enzyme (AHAS) complex. In addition, Rodrigues *et al.*²⁵ have found that *N*-benzoyl-*N'*-pyrimidin-2-yl thioureas experience π -stacking interactions involving the pyrimidinyl ring. Based on these outcomes, the unprecedented molecules **P1**–**P4** of Figure 7 were proposed and the respective pIC_{50} could be calculated using Eqs. (1) and (2). Accordingly, 3 out the 4 proposed compounds presented predicted pIC_{50} values higher than those of the library compounds, reinforcing the important role of the selected substituents to the *B. napus* growth inhibition and their synergistic effect.

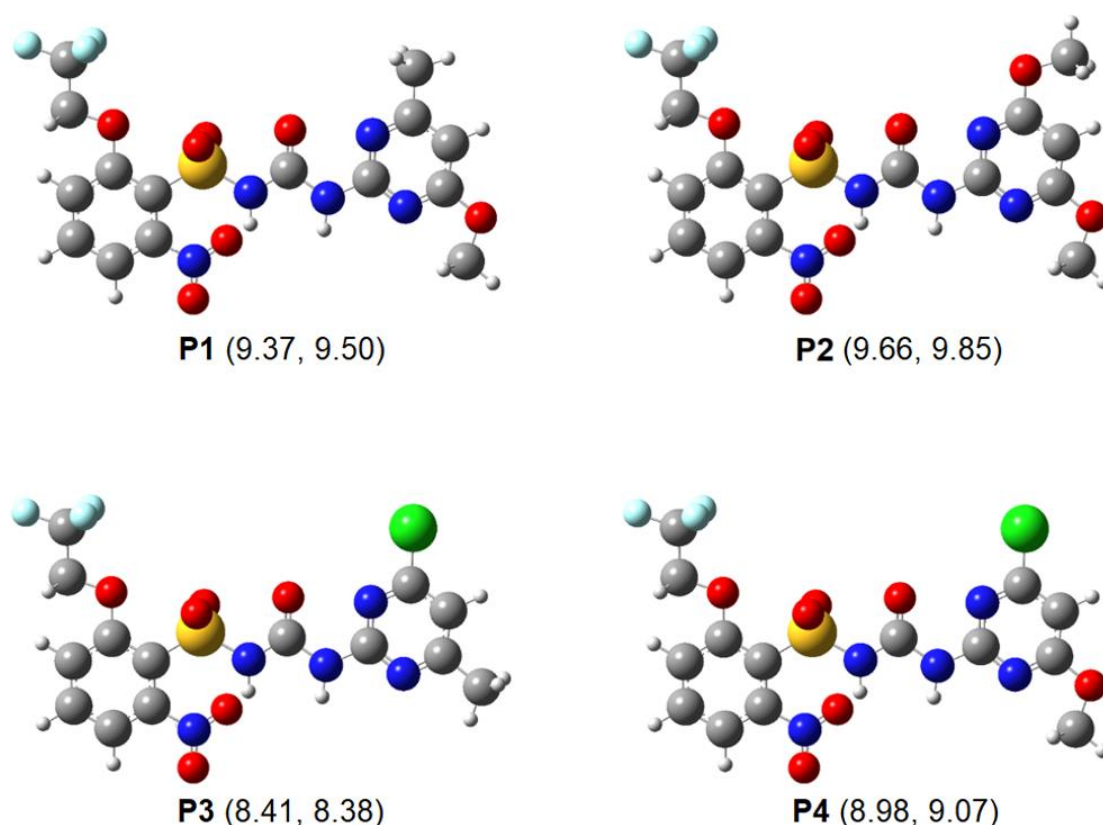


Figure 7. Proposed compounds **P1**–**P4** and the corresponding calculated values of pIC_{50} ,

obtained from MIA_{rvdW} and MIA_{ϵ} descriptors, respectively. H = white, C = gray, O = red, N = blue, F = cyan, S = yellow, Cl = green.

It is worth mentioning that experimental assays should be performed to attest the theoretical findings. In addition, caution should be taken when interpreting the biological response, since the decay of sulfonylureas into sulfonamide and other products on the pyrimidine side may originate more or less active compounds, thus interfering in the observed outcome. For example, inductive effects (-M or +M) of aryl substituents may affect the stability of the linkage symbolized by -||- in an Ar-SO₂-NH-||-CO-NH-pyrimidinyl ring compound. That is, electron-withdrawing groups, such as the NO₂ and OCH₂CF₃ groups in the proposed compounds rather than the chlorine substituent in chlorsulfuron, positivate the carbonyl carbon through Hammett sigma effects,²⁶ then facilitating the hydrolysis with nucleophilic water and leading to a bond cleavage and the decay of the compound into two halves, each bearing a ring.²

Conclusions

A few MIA descriptors were capable to model the inhibition data of 17 sulfonylurea herbicides. New derivatives were proposed based on the size and electron-withdrawing character of the substituents at various positions. The selected descriptors and MIA-contour maps revealed that pyrimidine analogs containing an OCH₂CF₃ group at the R¹ position are promising attempts to improve the potency of congeneric sulfonylureas and, most importantly, to tackle the problem of herbicide resistance by designing new chemicals.

Conflicts of interest

There are no conflicts do declare.

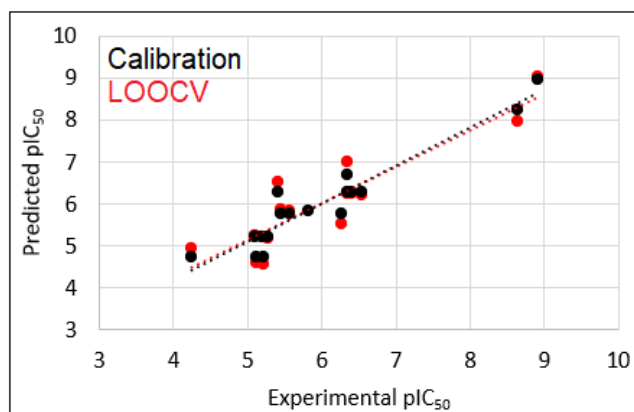
Acknowledgements

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 306830/2021-3), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

References

1. Ervin, D. E. Herbicide resistance: Economic and environmental challenges. *AgBioForum*, 2016, **19**, 136-155.
2. Paporisch, A.; Laor, Y.; Rubin, B. and Eizenberg, H. *Agronomy*, 2020, **10**, 1724.
3. Wei, W.; Zhou, S.; Cheng, D.; Li, Y.; Liu, J.; Xie, Y.; Li, Y. and Li, Z. *Bioorg. Med. Chem. Lett.*, 2017, **27**, 3365-3369.
4. Qu, R.-Y.; He, B.; Yang, J.-F.; Lin, H.-Y.; Yang, W.-C.; Wu, Q.-Y.; Li, Q. X. and Yang, G.-F. *Pest Manag. Sci.*, 2021, **77**, 2620-2625.
5. Finney, D. J. *J. Inst. Actuar.*, 1952, **78**, 388-390.
6. Surhone, L. M.; Timpledon, M. T. and Marseken, S. F. *NCSS (Statistical Software)*. Betascript Publishing: Mauritius, 2010.
7. Levitt, G. *American Chemical Society Symposium Series 443*. Washington D.C.; 1991, 16-31.
8. Todeschini, R.; Consonni, V.; Ballabio, D. and Grisoni, F. Chemometrics for QSAR Modeling, In *Comprehensive Chemometrics*. Elsevier: Amsterdam, 2020.
9. Freitas, M. R.; Barygie, S. J. and Freitas, M. P. *RSC Adv.*, 2015, **5**, 7547-7553.
10. Dennington, R. D.; Keith, T. A. and Millam, M. J. GaussView 5.0, Wallingford, CT, 2008.
11. Ambure, P.; Aher, R. B.; Gajewicz, A.; Puzyn, T. and Roy, K. *Chemometr. Intell. Lab. Syst.*, 2015, **147**, 1-13.
12. Kennard, R. W. and Stone, L. A. *Technometrics*, 1969, **11**, 137-148.
13. Tropsha, A. *Mol. Inf.*, 2010, **29**, 476-488.
14. Geladi, P. and Kowalski, B. R. *Anal. Chim. Acta*, 1986, **185**, 1-17.
15. Freitas, M. R.; Barigye, S. J. and Freitas, M. P. *RSC Adv.*, 2015, **5**, 7547-7553.
16. de Jong, S. *Chemom. Intell. Lab. Syst.*, 1993, **18**, 251-263.
17. Farrés, M.; Platikanov, S.; Tsakovski, S. and Tauler, R. *J. Chemom.*, 2015, **29**, 528-536.
18. Nunes, C. A.; Freitas, M. P.; Pinheiro, A. C. M. and Bastos, S. C. *J. Braz. Chem. Soc.*, 2012, **23**, 2003-2010.
19. V. Consonni, D. Ballabio and R. Todeschini, *J. Chem. Inf. Model.*, 2009, **49**, 1669-1678.
20. K. Roy, P. Chakraborty, I. Mitra, P. K. Ojha, S. Kar and R. N. Das, *J. Comput. Chem.*, 2013, **34**, 1071-1082.

21. Cramer, R. D.; Patterson, D. E. and Bunce, J. D., *J. Am. Chem. Soc.*, 1988, **110**, 5959-5967.
22. I. Mitra, A. Saha and K. Roy, *Mol. Simul.*, 2010, **36**, 1067-1079.
23. J.-H. Li, Y. Wang, Y.-P. Wu, R.-H. Li, S. Liang, J. Zhang, Y.-G. Zhu and B.-J. Xie, *Pest. Biochem. Physiol.*, 2021, **172**, 104766.
24. J.-H. Li, R.-H. Li, Y. Wang, S.-X. Li, Y.-P. Wu, J. Zhang, Y.-G. Zhu and B.-J. Xie, *Phosphorus Sulfur Silicon Relat. Elem.*, 2021, **196**, 664-671.
25. N. E. Rodrigues, A. C. de Faria, I. V. Pereira, E. F. F. da Cunha and M. P. Freitas, *Bull. Environ. Contam. Toxicol.*, 2022, <https://doi.org/10.1007/s00128-022-03467-w>.
26. Yadav, V. K., Hammett Substituent Constants. In: *Steric and Stereoelectronic Effects in Organic Chemistry*. Springer: Cham, 2021.



GA/MLR



ARTIGO 3

Artigo submetido na revista: Soil and Sediment Contamination An International Journal

Chemical structural features affecting the soil sorption of known sulfonylurea herbicides

The contamination of groundwater by herbicides depends on their mobility in soil, which is related to the organic carbon to water partition coefficient, K_{OC} . Because the log K_{OC} values of widely used sulfonylurea herbicides do not correlate with easily accessible log P estimations, multivariate image analysis (MIA) descriptors were employed to build structure-property models to classify these compounds according to their soil sorption capability. In addition, a deeper analysis based on MIA contour maps of PLS regression coefficients and variable importance in projection scores was performed to obtain insight into the chemical features responsible for the log K_{OC} behavior. A multiple linear regression model obtained from selected descriptors demonstrated high predictability ($r^2 = 0.95$, $q^2 = 0.84$, and $r^2_{pred} = 0.71$) and also provided chemical insight to the log K_{OC} values. It has been found that triazine rather than pyrimidine derivatives are less prone to leach out, but substituents attached to the sulfonyl group play an important role in modulating the log K_{OC} values.

Keywords: structure-property; log K_{OC} ; sulfonylureas; leaching; multivariate image analysis

Introduction

Sulfonylurea compounds, which are acetohydroxyacid synthase (AHAS) inhibitors, have been increasingly used as herbicides for weed control in cereals (Pannell et al., 2016). They can be divided into the pyrimidinylsulfonylureas and the triazinylsulfonylureas, and are taken up by the roots and foliage and act by disrupting protein synthesis (Sarmah and Sabadie, 2002). Although these herbicides are generally only slightly toxic to freshwater fish and invertebrates, and practically nontoxic to wildfowl and other mammals (Kamrin, 1997), their behavior in the

environment should be better understood to aid the design of safer analogues. For example, degradation of sulfonylureas in soil is pH-dependent, while the effects of side products obtained from their decay are not completely understood, that is, sulfonamide and other by-products may be more or less active, and more or less toxic compounds than the original sulfonylurea (Paporisch et al., 2020). In neutral- to low-alkaline soils, degradation of sulfonylureas by chemical hydrolysis is slow and, therefore, those more lipophilic, with higher pK_a , and less soluble in water would persist in soils rich in organic matter to a higher extent (Paporisch et al., 2020).

An important environmental indicator is the organic carbon to water partition coefficient, K_{OC} , defined as:

$$K_{OC} = (K_D \times 100)/f_{OC} \quad (1)$$

where the distribution coefficient K_D corresponds to the ratio C_S/C_W (C_S = concentration of herbicide adsorbed in the soil, and C_W = concentration of herbicide in the soil solution), and f_{OC} is the content of organic carbon in the soil (%).

The larger the K_{OC} , the higher the affinity of an herbicide to the organic fraction of the soil, although the presence of clay is one of the factors conditioning this process. Therefore, herbicides with high mobility in soil (low K_{OC}) can leach to groundwater and affect wild aquatic plant growth (Wauchope et al., 1992). Consequently, understanding the chemical features affecting K_{OC} can aid the design of novel eco-friendly sulfonylurea herbicides. Although the octanol-water partition coefficient $\log P$ has been demonstrated to correlate to some extent with the soil sorption of a variety of herbicides (Sabljic et al., 1995), this does not always apply in other cases (Freitas et al., 2014a, 2014b, 2015). In this sense, more complex modeling strategies may be required to accurately estimate the behavior of K_{OC} of interesting herbicides. Numerous sulfonylureas are currently registered with the U.S. EPA, and 17 are measured for the *National Report on Human Exposure to Environmental Chemicals* (https://www.cdc.gov/biomonitoring/SulfonylureaHerbicides_BiomonitoringSummary.html). Of these compounds, to the best of our knowledge, only halosulfuron does not have a measured K_{OC} value and, therefore, the remaining 16 sulfonylureas (Figure 1) were used to obtain pattern recognition and interpretative models using multivariate image analysis (MIA) molecular descriptors (Barigye and Freitas, 2016).

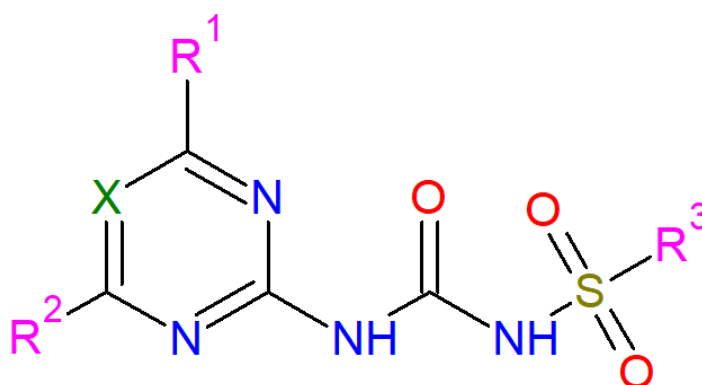


Figure 1- Basic structure of the pyrimidinylsulfonylureas and triazinylsulfonylureas studied in this work.

MIA descriptors are based on RGB (red-green-blue) pixel values of images corresponding to colored chemical structures; the colors assigned to each atom type can encode atomic properties, such as van der Waals radius (r_{vdW}) and Pauling's electronegativity (ϵ). In addition, these descriptors may be regressed against the response data ($\log K_{OC}$) using partial least squares (PLS) to give variable importance in projection (VIP) scores and PLS regression coefficients (**b**), which inform on how much and how (increasing or decreasing) the MIA descriptors affect the dependent variable (Barigye et al., 2016). PLS is a multivariate regression method that finds a linear correlation by projecting the predicted variables and the observable variables to a new space, called latent variables (Geladi and Kowalski, 1986). The PLS model is a linear combination of the original independent variables (MIA descriptors) weighted by the vector of regression coefficients **b** (Farrés et al., 2015). The VIP scores are the weighted sum of squares of the weights, which measures the percentage of the variance in the $\log K_{OC}$ values explained with each latent variable (de Jong, 1993).

Materials and Methods

The $\log K_{OC}$ values for 16 sulfonylureas measured for the *National Report on Human Exposure to Environmental Chemicals* were obtained from the PubChem database and were mostly reported by the U.S. EPA Distributed Structure-Searchable Toxicity. Standard pH 7 at 20°C is implicit. The molecular descriptors used for the structure-activity relationship (SAR) analysis were based on multivariate image analysis (MIA), whose modeling procedure is described as follows. The molecular structures were generated using the GaussView program (Dennington et al., 2008), using the ball-and-stick scheme and atoms sized according to the respective van der Waals radii. The superposable 2D chemical structures were converted to bitmap images of

300 × 354 pixels (Figure 2). The pixel values were renumbered to make a correspondence with Pauling's electronegativity (MIA- ϵ model), van der Waals radius (MIA- r_{vdW} model), and r_{vdW}/ϵ ratio (MIA- r_{vdW}/ϵ model). The images were joined to form a three-way array of dimension 16 × 300 × 354, which was then unfolded to a 16 × (300 × 354) matrix. This **X**-matrix was submitted to principal component analysis (PCA) without any data pre-processing using the Chemoface program (Nunes et al., 2012). In addition, the **X**-matrix was regressed against the log K_{OC} values using partial least squares (PLS) regression and 2 latent variables to obtain the PLS regression coefficients (**b**) and variable importance in projection (VIP) scores. These parameters were used to build the MIA plots (Barigye et al., 2016). A MIA-QSAR model using electronegativity-based descriptors selected by genetic algorithm (GA) was built, and these descriptors were regressed against the log K_{OC} values using multiple linear regression (MLR). The fitness function employed in the GA selection was based on error/MAE-based parameters, and a total of 100 iterations, 0.3 mutation probability, and 30 equations selected in each generation were applied (Ambure et al., 2015). The model's quality was determined by the determination coefficient of calibration (r^2), leave-one-out cross-validation (q^2), and prediction (r^2_{pred}).

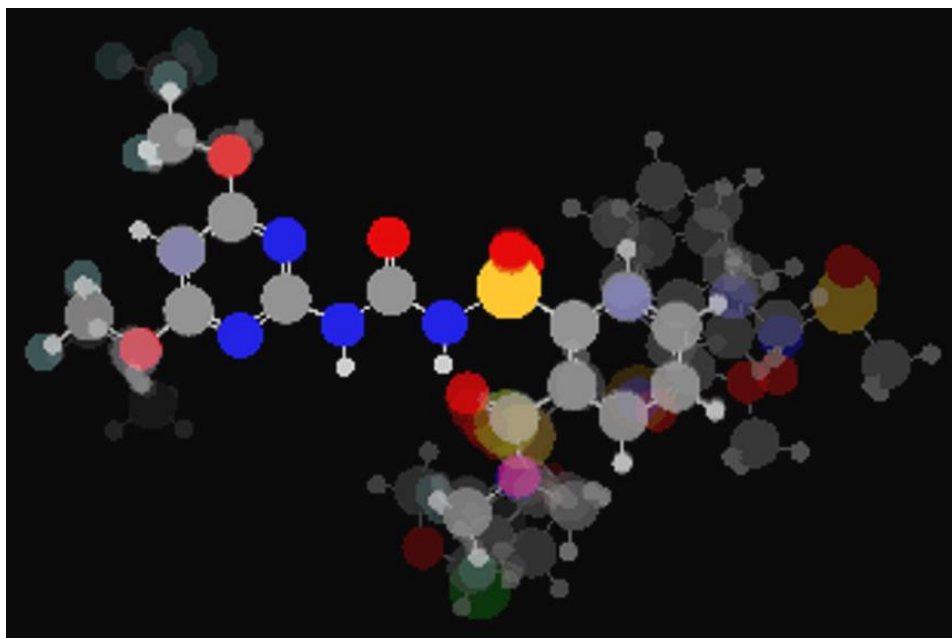


Figure 2- Superposed images of the 16 sulfonylurea herbicides studied. Notice the congruent (congeneric) and variable (substituents) moieties responsible for the soil sorption data.

Results and Discussion

Since the log K_{OC} for halosulfuron within the series of 17 sulfonylureas was not available, a preliminary pattern recognition study was performed using principal component analysis (PCA) of the MIA descriptors for the remaining 16 widely used sulfonylurea herbicides (Figure 1 and Table 1).

Table 1. Sulfonylurea herbicides used in the structure-property study of soil sorptions (log K_{OC}).

No.	Name	X	R ¹	R ²	R ³	log K_{OC}
1	Bensulfuron methyl	CH	OMe	OMe	-CH ₂ (2-COOMe)Ph	2.50
2	Foramsulfuron	CH	OMe	OMe	-(2-CONMe ₂ ,5-NHCHO)Ph	1.00
3	Mesosulfuron methyl	CH	OMe	OMe	-(2-COOMe,5-CH ₂ NHSO ₂ Me)Ph	1.04
4	Nicosulfuron	CH	OMe	OMe	-2(3-CONMe ₂)Pyr	1.48
5	Rimsulfuron	CH	OMe	OMe	-2(3-SO ₂ Et)Pyr	1.71
6	Sulfosulfuron	CH	OMe	OMe	-3(2-(SO ₂ Et)imidazo[1,2- <i>a</i>]pyridine)	1.52
7	Primisulfuron methyl	CH	OCF ₃	OCF ₃	-(2-COOMe)Ph	1.08
8	Sulfometuron methyl	CH	Me	Me	-(2-COOMe)Ph	1.62
9	Oxasulfuron	CH	Me	Me	-(2-COOC ₃ H ₆ O)Ph	1.93
10	Chlorsulfuron	N	OMe	Me	-(2-Cl)Ph	2.19
11	Metsulfuron methyl	N	OMe	Me	-(2-COOMe)Ph	1.54
12	Prosulfuron	N	OMe	Me	-(2-CH ₂ CH ₂ CF ₃)Ph	1.15
13	Triasulfuron	N	OMe	Me	-(2-OCH ₂ CH ₂ Cl)Ph	2.02
14	Thifensulfuron methyl	N	OMe	Me	-3(2-COOMe)thiophene	1.45
15	Ethametsulfuron methyl	N	OEt	NHMe	-(2-COOMe)Ph	1.93
16	Triflusulfuron methyl	N	OCH ₂ CF ₃	NMe ₂	-(2-COOMe,6-Me)Ph	1.84

^a Halosulfuron was removed from the data set due to a lack of information.

The first attempt to classify the sulfonylurea compounds using the three most important principal components, which explain 80.48% of data variance, generated three clusters of molecules in a 3D-space using MIA- ϵ descriptors: 1) **1**, **8**, **9**; 2) **2** to **7**; and 3) **10** to **16** (Figure 3a). Careful analysis of this clustering allowed us to assess the following: the first cluster is formed by pyrimidinylsulfonylureas containing unique substituents at the R¹, R², and R³ positions, *i.e.* -CH₂(2-COOMe)Ph at R³ for **1**, and (3,5-diMe)Ph at R¹ and R² for **8** and **9**. The second cluster (**2** to **7**) includes pyrimidine derivatives with OMe groups at R¹ and R² (OCF₃ for **7**), while the third cluster contains the triazinylsulfonylureas. The PCA score plots obtained from the MIA-r_{vdw} and MIA-r_{vdw}/ ϵ descriptors (Figures 3b and 3c) were similar to each other and the same conclusion can be drawn from both analyses, *i.e.* PC3 discriminates the pyrimidinyl and triazinylsulfonylureas, whereas PC2 separates compounds **1** and **6** from the others, because of the structurally diverse groups at R³ (-CH₂Ph in **1** and imidazo[1,2-a]pyridine in **6**) in these compounds. In turn, the PC1 \times PC2 plot obtained from all three MIA descriptor types separates reasonably well the compounds with low and high log K_{OC} values (Figure 3d). Considering that the log K_{OC} values span from 1.00 to 2.50 along with this series of compounds, the superior half (log K_{OC} > 1.75) is mostly represented by triazine-containing scaffolds. That is, the X group (N or CH) plays an important role in describing the soil sorption of a sulfonylurea herbicide in soil and, therefore, pyrimidine rather than triazine derivatives are more prone to leach out and contaminate the groundwater.

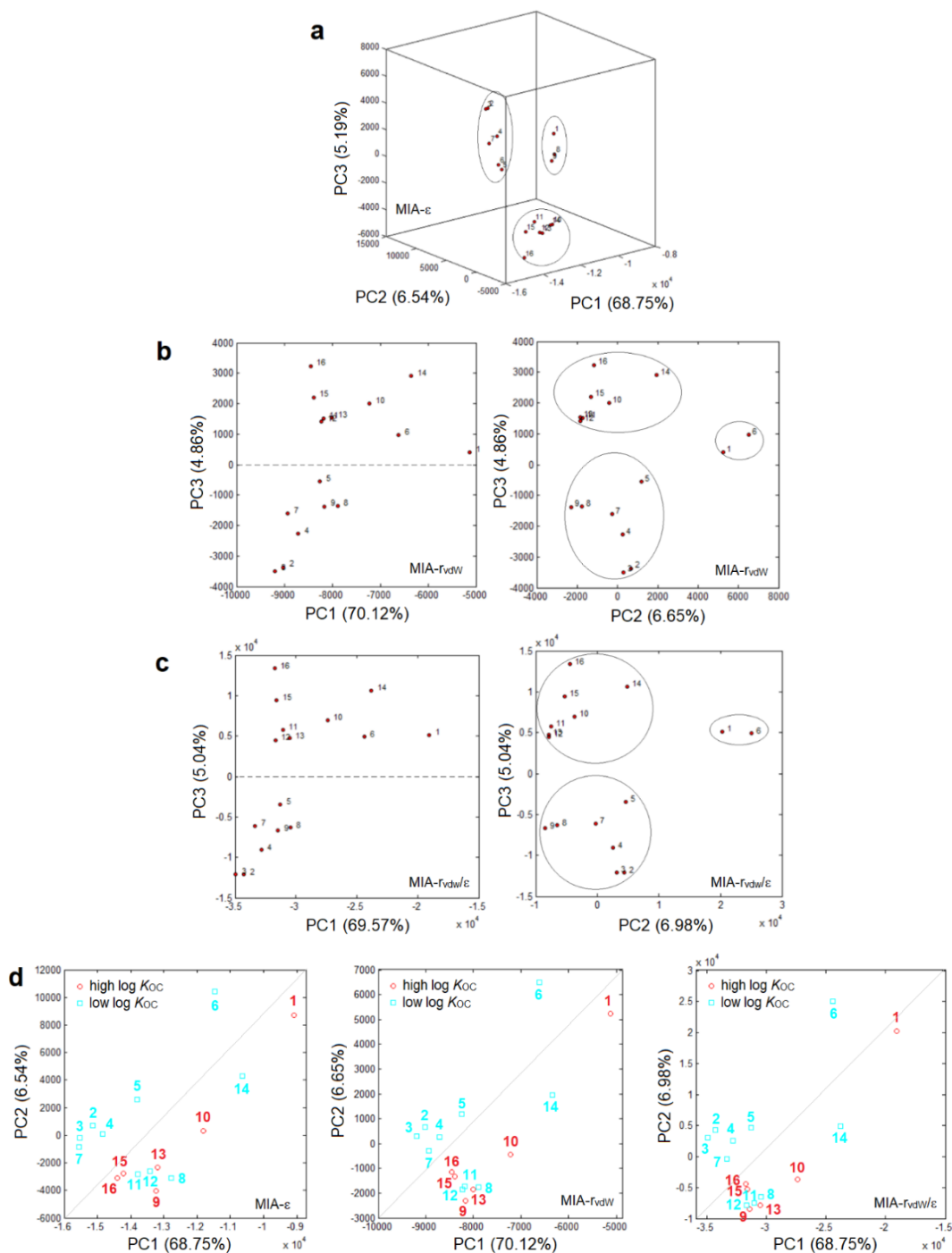


Figure 3- a) Three-way PCA scores using MIA- ϵ descriptors showing the clustering of 3 groups: compounds **1, 8, 9**; compounds **2 to 7** (pyrimidine derivatives); compounds **10 to 16** (triazine derivatives); b) two-way PCA score plots using MIA- r_{vdw} descriptors showing 2 clusters in PC3 (PC1 \times PC3 plot) and 3 clusters in the PC2 \times PC3 plot; c) two-way PCA score plots using MIA- $r_{vdw/\epsilon}$ descriptors showing 2 clusters in PC3 (PC1 \times PC3 plot) and 3 clusters in the PC2 \times PC3 plot; d) two-way PCA score plots in PC1 \times PC2 separating compounds with low (< 1.75) and high (> 1.75) $\log K_{oc}$ values. The percentages in parenthesis indicate the explained variance by each principal component.

To gain detailed insights into the substituents (R^1 , R^2 , and R^3) affecting the $\log K_{OC}$ values, an MIA-plot analysis considering the PLS regression coefficients (**b**) and variable importance in projection (VIP) scores was performed. It is worth mentioning that the data set could be satisfactorily calibrated, as the determination coefficient R^2 in all three models was *ca.* 0.8 (Figure 4). The MIA plots (Figure 5) obtained from MIA descriptors encoding ε , r_{vdW} , and r_{vdW}/ε properties are very similar and, therefore, a unified analysis is provided herein. According to the VIP plot, substituents attached to the aromatic ring at R^3 demonstrate a strong effect on $\log K_{OC}$. Atomic contributions come particularly from oxygen and nitrogen atoms of the ester and amide groups at the 2-position of the aromatic ring in R^3 , as well as the amide and sulfonamide nitrogens at the 5-position. According to the **b**-plot, these contributions are negative, *i.e.*, the above-mentioned groups tend to decrease $\log K_{OC}$. In turn, the chlorine atom in **11** and **14**, the oxetane ring in **9**, and the substituted $-CH_2Ph$ moiety in **1** at R^3 contribute to increasing the $\log K_{OC}$.

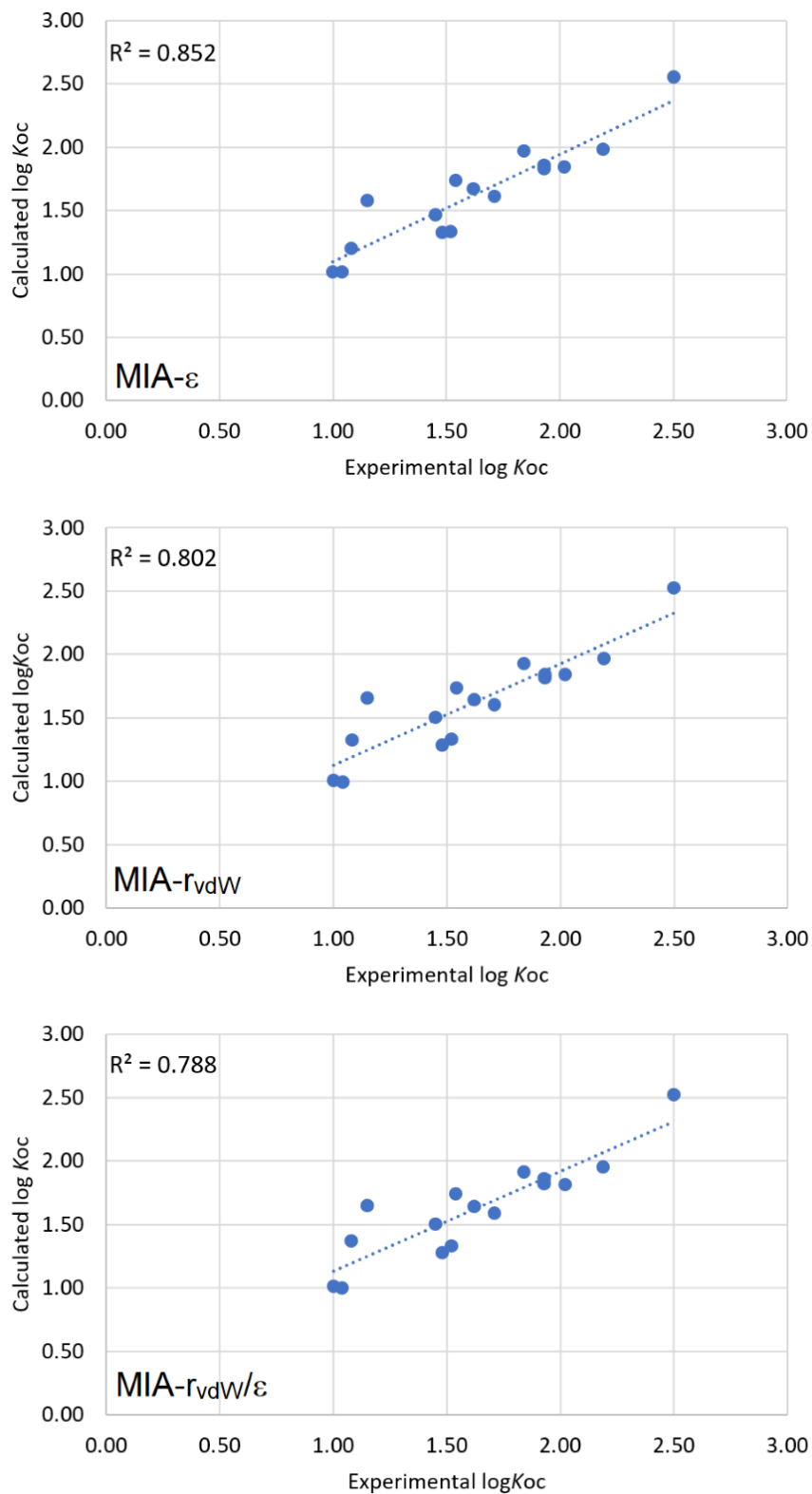


Figure 4- Correlation plots of experimental \times calculated $\log K_{oc}$ using MIA descriptors.

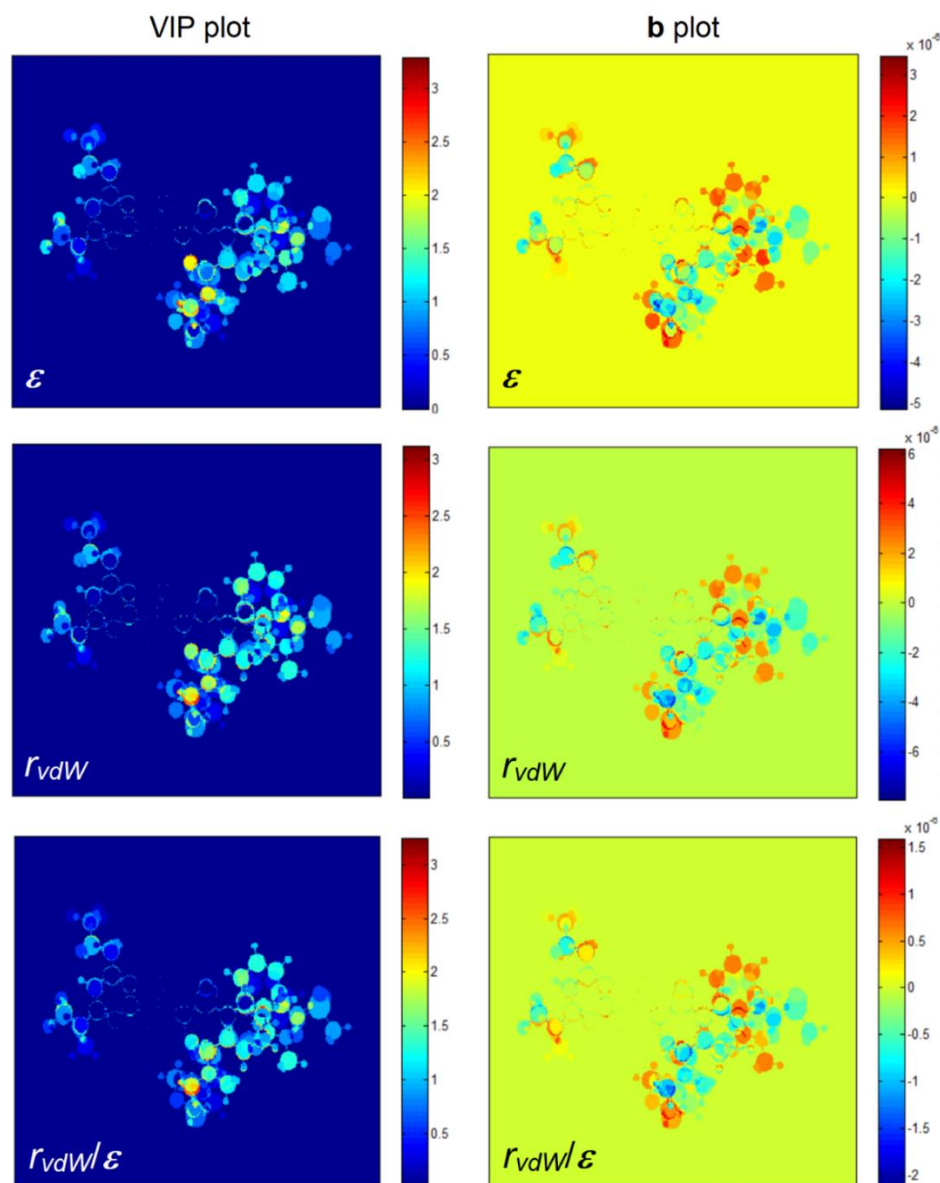


Figure 5- MIA plots indicating how (increasing or decreasing, **b**-plot) and how much (VIP-plot) the MIA descriptors influence the log K_{OC} values for the 16 sulfonylurea herbicides.

Although the partial least squares regression for the whole set of MIA descriptors against the log K_{OC} values did not yield satisfactory results in leave-one-out cross-validation ($q^2 < 0.5$) and external validation ($r^2_{pred} < 0.5$), a few MIA_{ϵ} descriptors selected by genetic algorithm (Table 2) demonstrated high performance in both calibration and validation (Figure 6). The corresponding multiple linear regression model (Eq. 1) indicates that absence of atom (px = 0 in **8** and **9** due to a small Me group that does not occupy that pixel orientation R^1), and the presence of C (px = 255 in **3** and **6**), S (px = 258 in **6** due to the SO_2Et group), N (px = 304 in **2** due to the $NHCHO$ group), and F (px = 398 in **7** due to the CF_3 group in R^1) atoms in certain image coordinates contribute importantly to the model. It is worth noticing from Eq. 1 that the four selected molecular descriptors (MIA_{a-d} pixels – px) have a negative influence on

$\log K_{OC}$, that is, larger values (electronegativity) for each descriptor contribute to decreasing the response variable. This is reasonable, as polar compounds tend to be more mobile in soils rich in organic matter.

$$\log K_{OC} = 2.725 - (0.0017 \times MIA_a) - (0.0019 \times MIA_b) - (0.0018 \times MIA_c) - (0.0014 \times MIA_d)$$

(Eq. 1)

Table 2. MIA descriptors proportional to electronegativity values selected by genetic algorithm for use in the multiple linear regression.

Cpd	MIA _a	MIA _b	MIA _c	MIA _d
1	220	0	0	0
2	220	255	255	304
3	220	255	255	255
4	220	255	255	0
5	220	255	0	0
6	220	0	258	255
7	398	255	255	0
8	0	255	255	0
9	0	255	255	0
10	220	0	0	0
11	220	255	255	0
12	220	255	255	0
13	220	255	0	0
14	220	255	255	0
15	220	0	255	0
16	220	0	255	0

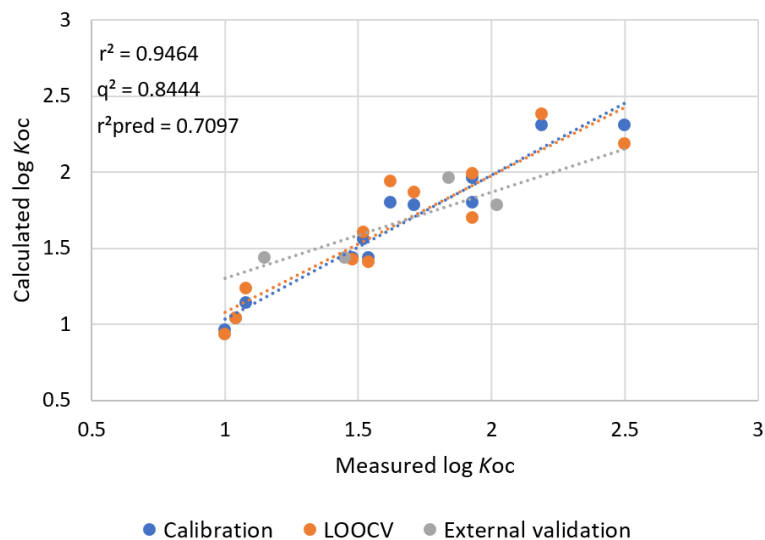


Figure 6- Plot of measured \times calculated log K_{OC} obtained by the MLR model using MIA descriptors based on electronegativity values. Test samples selected by Kennard-Stone sampling: **12, 13, 14, and 16**.

In summary, triazinylsulfonylureas are in general more persistent in soil than pyrimidine analogues, which in turn have higher mobility and more tendency to leach out. Secondary effects stemming from substituents at R^3 also modulate the log K_{OC} values, and this can be an important point to tackle when designing novel sulfonylurea herbicides. MIA descriptors properly described the soil sorption of an important series of compounds, whereas the lipophilicity parameter log P did not correlate with log K_{OC} (Figure 7).

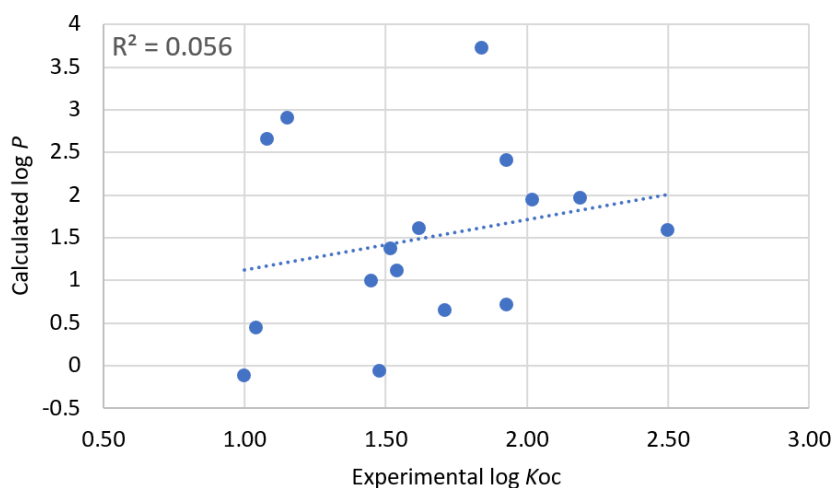


Figure 7- Plot of log $K_{OC} \times$ log P (calculated in <https://www.molinspiration.com/>) showing the lack of correlation between soil sorption and lipophilicity for the 16 sulfonylurea herbicides.

As a concluding remark, based on the SAR and QSAR outcomes, we suggest triazinylsulfonyleureas with hydrophobic substituents at the R³ position, such as the chlorine substituent, as the best choice for further synthesis.

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 306830/2021-3), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

The authors declare no conflict of interest.

References

- Ambure, P., Aher, R.B., Gajewicz, A., Puzyn, T., Roy, K. 2015. NanoBRIDGES” software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemometr. Intell. Lab. Syst.* **147**, 1-13. <https://doi.org/10.1016/j.chemolab.2015.07.007>.
- Barigye, S.J., Duarte, M.H., Nunes, C.A., Freitas, M.P. 2016. MIA-plot: A graphical tool for viewing descriptor contributions in MIA-QSAR. *RSC Adv.* **6**, 49604-49612. <https://doi.org/10.1039/C6RA09593C>.
- Barigye, S.J., Freitas, M.P. 2016. Ten years of the MIA-QSAR strategy: Historical development and applications. *Int. J. Quant. Struct. Prop. Relat.* **1**, 64-77. <https://doi.org/10.4018/IJQSPR.2016010103>.
- de Jong, S. 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemometr. Intell. Lab. Syst.* **18**, 251-263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X).
- Dennington, R.D., Keith, T.A., Millam, M.J. 2008. GaussView 5.0, Wallingford, CT.
- Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R. 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemometr.* **29**, 528-536. <https://doi.org/10.1002/cem.2736>.

- Freitas, M.R.; Matias, S.V.B.G., Macedo, R.L.G., Freitas, M.P., Venturin, N. 2014a. Three-parameter modeling of the soil sorption of acetanilide and triazine herbicide derivatives. *Bull. Environ. Cont. Toxicol.* **92**, 143-147. <https://doi.org/10.1007/s00128-013-1184-3>.
- Freitas, M.R., Freitas, M.P., Macedo, R.L.G. 2014b. Aug-MIA-QSPR modeling of the soil sorption of carboxylic acid herbicides. *Bull. Environ. Cont. Toxicol.* **93**, 489-492. <https://doi.org/10.1007/s00128-014-1356-9>.
- Freitas, M.R., Barigye, S.J., Freitas, M.P. 2015. Coloured chemical image-based models for the prediction of soil sorption of herbicides. *RSC. Adv.* **5**, 7547-7553. <https://doi.org/10.1039/C4RA12070A>.
- Geladi, P., Kowalski, B. R. 1986. Partial least squares regression: A tutorial. *Anal. Chim. Acta* **185**, 1-17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- Kamrin, M., 1997. *Pesticide Profiles-Toxicity, Environmental Impact, and Fact*. CRC-Lewis Publishers, Boca Raton.
- Nunes, C.A., Freitas, M.P., Pinheiro, A.C.M., Bastos, S.C. 2012. Chemoface: A novel free user-friendly interface for chemometrics. *J. Braz. Chem. Soc.* **23**, 2003-2010. <https://doi.org/10.1590/S0103-50532012005000073>.
- Pannell, D.J., Tillie, P., Rodríguez-Cerezo, E., Ervin, D., Frisvold, G.B. 2016. Herbicide resistance: economic and environmental challenges. *AgBioForum* **19**, 141-155.
- Paporisch, A., Laor, Y., Rubin, B., Eizenberg, H. 2020. Effect of repeated application of sulfonylurea herbicides on sulfosulfuron dissipation rate in soil. *Agronomy* **10**, 1724. <https://doi.org/10.3390/agronomy10111724>.
- Sabljić, A., Güsten, H., Verhaar, H., Hermens, J. 1995. QSAR modelling of soil sorption. Improvements and systematics of logK_{OC} versus logK_{OW} correlations. *Chemosphere* **31**, 4489-4514. [https://doi.org/10.1016/0045-6535\(95\)00327-5](https://doi.org/10.1016/0045-6535(95)00327-5).
- Sarmah, A.K., Sabadie, J. 2002. Hydrolysis of sulfonylurea herbicides in soils and aqueous solutions: a review. *J. Agric. Food Chem.* **50**, 6253-6265. <https://doi.org/10.1021/jf025575p>.
- Wauchope, R.D., Buttler, T.M., Hornsby, A.G., Augustijn-Beckers, P.W., Burt, J.P. 1992. The SCS/ARS/CES pesticide properties database for environmental decision-making. *Rev. Environ. Contam. Toxicol.* **123**, 1-155. https://doi.org/10.1007/978-1-4612-2862-2_1.