



INGRID VITTORIA PEREIRA

**MODELAGEM QSAR DAS ATIVIDADES HERBICIDAS DE
BENZAMIDAS PARA CONTROLE DE PRAGAS NA
AGRICULTURA**

**LAVRAS - MG
2022**

INGRID VITTORIA PEREIRA

**MODELAGEM QSAR DAS ATIVIDADES HERBICIDAS DE
BENZAMIDAS PARA CONTROLE DE PRAGAS NA
AGRICULTURA**

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

Prof. Dr. Matheus Puggina de Freitas
Orientador

**LAVRAS - MG
2022**

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

Pereira, Ingrid Vittoria.

Modelagem QSAR das atividades herbicidas de Benzamidas
para controle de pragas em agricultura / Ingrid Vittoria Pereira. -
2022.

79 p. : il.

Orientador(a): Matheus Puggina de Freitas.

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

Bibliografia.

1. QSAR. 2. Herbicida. 3. Fotossistema II. I. Freitas, Matheus
Puggina de. II. Título.

INGRID VITTORIA PEREIRA

**QSAR MODELING OF BENZAMIDE HERBICIDAL ACTIVITIES FOR PEST
CONTROL IN AGRICULTURE**

**MODELAGEM QSAR DAS ATIVIDADES HERBICIDAS DE BENZAMIDAS PARA
CONTROLE DE PRAGAS NA AGRICULTURA**

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

APROVADA EM 9 de fevereiro de 2022.

Dra. Daiana Teixeira Mancini
Profa. Dra. Elaine Fontes Ferreira da Cunha

InSILICO
UFLA

Prof. Dr. Matheus Puggina de Freitas
Orientador

**LAVRAS – MG
2022**

*Aos meus, pais Mário e Marcelena, por terem abdicado de suas vidas em prol das realizações
e felicidade de suas filhas.*

A minha irmã Yasmim, por sua preocupação e todo amor e incentivo.

As minhas avós Maria e Lídia.

*Aos demais membros de minha família, por se alegrarem nas minhas vitórias e me apoiarem
nas derrotas.*

A Deus, pela vida, saúde e por nunca me abandonar.

Dedico!!

AGRADECIMENTO

Agradeço primeiramente a Deus, por me permitir a existência, por direcionar minha vida e por permitir que mais essa etapa na minha vida fosse cumprida, tendo a certeza que sempre esteve ao meu lado em todos os momentos.

Ao professor Matheus Puggina de Freitas, pela paciência, atenção, confiança e orientação, por ter me aberto as portas à pesquisa, mesmo diante de minhas dificuldades, para que pudéssemos concretizar este trabalho, meus sinceros agradecimentos.

A professora Elaine Fontes Ferreira da Cunha pela contribuição nos estudos de ancoramento molecular.

À banca examinadora pela disponibilidade e contribuições para este trabalho.

Ao Grupo de Pesquisa do Laboratório de Química Computacional da UFLA – MOLECC – UFLA, que me acompanhou durante esses dois anos e por toda amizade construída, sempre estarão na minha memória e coração, em especial à Joyce Karoline Daré, por toda ajuda durante essa caminhada.

Aos meus pais, Mário e Marcelena pelo carinho, incentivo e apoio em toda minha existência me ensinando a importância da educação, dedicação e trabalho.

A minha irmã, Yasmim por todo amor e torcida em importantes momentos da minha vida acadêmica.

A todos os professores e funcionários da UFLA.

À CAPES, FAPEMIG e CNPq, pelo apoio financeiro.

A todas as pessoas que, direta ou indiretamente, contribuíram para a execução dessa Dissertação de Mestrado.

MEU MUITO OBRIGADA!!!

“O sucesso é ir de fracasso em fracasso sem perder o entusiasmo”.

Ingrid Vittoria Pereira

RESUMO

É incontestável a necessidade do uso de herbicidas para o controle de plantas daninhas nos sistemas agrícolas. Os herbicidas benzamídicos consistem em uma classe de inibidores do sistema fotossintético II (FSII) amplamente utilizados para o controle deste tipo de problema. No entanto, o desenvolvimento de resistência por essas ervas daninhas aos herbicidas conhecidos, torna-se necessário uma busca contínua por novos agroquímicos. O estudo relatado aqui parte da combinação de duas séries congênicas de herbicidas (tio) benzamida em um único conjunto de dados e modelagem subsequente de suas atividades herbicidas contra FSII (pIC_{50}) utilizando análise de imagem multivariada aplicada à relação quantitativa estrutura-atividade (MIA-QSAR). Os modelos apresentaram-se robustos e preditivos (valores médios de $r^2 = 0,91$, $q^2 = 0,80$ e $r^2_{pred} = 0,81$) e foram utilizados para estimar o pIC_{50} de novos candidatos agroquímicos, os quais foram propostos com base na mistura química das subestruturas dos compostos mais ativos presentes no conjunto de dados. As características químicas que afetam à atividade herbicida foram analisadas usando mapas de contorno MIA, enquanto os valores pIC_{50} previstos foram comparados com pontuações de docking para apoiar os resultados MIA-QSAR. Os modelos MIA-QSAR, cujos descritores moleculares explicam informações topoquímicas, raio atômico e eletronegatividade de Pauling, apresentaram alto desempenho de modelagem. Consequentemente, os dados pIC_{50} para os candidatos a agroquímicos propostos foram calculados usando os parâmetros de regressão PLS dos modelos. O composto proposto possuindo uma porção tiobenzamida e cadeia C-11, H, NO_2 , OH e OH como substituintes variáveis foi a alternativa mais promissora.

Palavras-Chave: Estrutura-atividade; (tio)benzamidás; atividade herbicida; fotossistema II.

ABSTRACT

The need to use herbicides to control weeds in agricultural systems is incontestable. Benzamide herbicides consist of a class of photosynthetic system II (PSII) inhibitors widely used for weed control. However, the development of resistance by these weeds to the known herbicides requires an ongoing search for new agrochemicals. The study reported here starts from combining two congeneric series of (thio)benzamide herbicides into a single dataset and subsequent modeling of their herbicidal activities against PSII (pIC_{50}) using multivariate image analysis applied to quantitative structure-activity relationship (MIA-QSAR). The models were robust and predictive (mean values of $r^2 = 0.91$, $q^2 = 0.80$ and $r^2_{pred} = 0.81$) and were used to estimate the pIC_{50} of new agrochemical candidates, which were proposed based on the mixture chemistry of the substructures of the most active compounds present in the dataset. The chemical features affecting the herbicidal activities were analyzed using MIA contour maps, whereas the predicted pIC_{50} values were compared to docking scores to support the MIA-QSAR outcomes. The MIA-QSAR models, whose molecular descriptors explain topochemical information, atomic radius, and Pauling's electronegativity, presented high modeling performance. Accordingly, the pIC_{50} data for proposed agrochemical candidates were calculated using the PLS regression parameters of the models. The proposed compound possessing a thiobenzamide moiety and C-11 chain, H, NO_2 , OH, and OH as variable substituents was the most promising alternative.

Keywords: Structure-activity; (thio)benzamides; herbicidal activity; photosystem II.

LISTA DE FIGURAS

PRIMEIRA PARTE

Figura 1 - Estrutura química do herbicida pronamida, da classe das benzamidas.	20
Figura 2 - Estrutura química básica de uma benzamida, precursora dos herbicidas benzamídicos.	25
Figura 3 - Herbicidas inibidores do fotossistema II e suas classificações	26
Figura 4 - Visualização dos sítios de ligação de alguns herbicidas inibidores do FS II ao complexo proteico QB na membrana dos cloroplastos.....	27
Figura 5 - Estrutura do fotossistema II (FSII) complexado ao herbicida terbutryn.....	27
Figura 6 - Esquema utilizado para o tratamento das imagens e geração dos descritores MIA-QSAR.	31
Figura 7 - 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.	32
Figura 8 - MIA-Plots baseados nos parâmetros b e VIP para derivados de triazinas.....	33
Figura 9 - Imagens sobrepostas para uso na construção do Modelo 1 (imagem superior) e dos Modelos 2-4 (imagem inferior).	44

SEGUNDA PARTE

Figure 1. Commercial Pronamide herbicide and the benzamide derivatives studied herein.Changes in X (O or S) and R explain the variance in the herbicidal activity of this class of compound.....	56
Figure 2. Correlation plots of experimental \times calculated pIC50 for the four MIA-QSAR models.	61
Figure 3. MIA-plots based on variable importance in projection (VIP) scores and PLS regression coefficients (B) for the three augmented MIA-QSAR model.....	62
Figure 4. Important ligand-enzyme (PDB code: 3JCU) interactions obtained from the docking studies (the MolDock scores are given in the Supplementary Material).	63

Figure S1. Superposed images showing how the (thio)benzamide derivatives are represented and vary in a congeneric series according to the conventional and augmented MIA-QSAR approaches.	75
Figure S2. General workflow used to build the MIA-QSAR models.	75
Figure S3. Applicability domain tests for the four MIA-QSAR data sets. The point labels do not correspond to the compound numbering, since the test set samples vary with the descriptor type; samples which are out of the cut-off values correspond to compounds 14, 15, and 84 (the point label 59 at the border line in the MIA _{sketch} plot corresponds to 83).....	76
Figure S4. MolDock scores (kcal mol ⁻¹) for the ligands of the known and proposed benzamide herbicides. Highlights: red = seeming outliers (14 and 15); green = best experimental benzamide herbicide (56); yellow = most stable analog of 56 in the pocket (58); magenta = P1; gray = P2.	79

LISTA DE TABELAS

PRIMEIRA PARTE

- Tabela 1** - Compostos derivados de 3-nitro-2,4,6-trihidroxibenzamida, grupo alquicarbamatos salicilanilida e correspondentes valores de atividade inibitória do FSII.40
- Tabela 2** - Escala de cores RGB dos átomos obtidos no GaussView e os valores quimicamente relevantes pelos quais os valores de pixels foram substituídos. 46

SEGUNDA PARTE

- Table 1.** Statistical evaluation of the MIA-QSAR models using Kennard-Stone sampling to split the data set into training and test subsets. 60
- Table 2.** Proposed (thio)benzamide herbicides and the respective calculated pIC₅₀ values (mean values for the four MIA-QSAR models). 62
- Table S1.** Benzamide (X = O) and thiobenzamide (X = S) derivatives, and the respective pIC₅₀ values used in the MIA-QSAR modeling.^a72
- Table S2.** Original RGB pixel values and descriptor values proportional to atomic properties assigned to the atoms present in the molecules of the data set.74
- Table S3.** Statistical results of robustness and stability obtained from the bootstrapping experiments (mean values for 20 cycles). 77
- Table S4.** Predicted pIC₅₀ values for the proposed compounds P1–P18 using the four MIA-QSAR models. 78

LISTA DE QUADROS

PRIMEIRA PARTE

QUADRO 1. - Classificação das metodologias QSAR, com base na dimensionalidade do descritor.....	29
---	----

LISTA DE ABREVIATURAS E SIGLAS

ALS	Acetolactato sintase
Aug-MIA-QSAR	Augmented Multivariate Image Analysis applied to <i>Quantitative Structure-Activity Relationship</i>
ANN	Redes Neurais Artificias
b-plot	Gráfico de coeficientes de regressão do PLS
Bmp	Bimaps
Ca	Cálcio
CCC	<i>Concordance Correlation Coefficient</i>
DA	Domínio de aplicabilidade
Fe	Ferro
FS II	Fotossistema II
LOOCV	<i>leave-one-out cross-validation</i>
LV	Variáveis Latentes
MAE	<i>Mean Absolute Error</i>
MIA	<i>Multivariate Image Analysis</i>
MIA-plot	Gráficos MIA
MIA-QSAR	<i>Multivariate Image Analysis applied to Quantitative Structure-Activity Relationship</i>
MLR	Regressão Linear Múltipla
Mn	Manganês
PLS	Mínimos Quadrados Parciais
r^2	Coeficiente de determinação
r^2_c	Coeficiente de determinação da calibração
r^2_m	Coeficiente de determinação modificado
r^2_{rand}	Coeficiente de determinação Y randomization
r^2_{test}	Coeficiente de determinação da validação externa
RMSE	<i>root mean square error</i>
q^2	Coeficiente de determinação da validação cruzada
$Q^2_{F1}, Q^2_{F2}, Q^2_{F3}$	<i>Predictive Squared Correlation Coefficients</i>
QSAR	<i>Quantitative Structure-Activity Relationship</i>
RGB	Red, Green, and Blue

PDB	<i>Protein Data Bank</i>
PET	Transporte de elétrons fotossintéticos
pIC ₅₀	Logaritmo negativo da concentração inibitória
VIP	Gráfico de <i>scores</i> da importância das variáveis em projeção
1D	Unidimensional
2D	Bidimensional
3D	Tridimensional
4D	Quarta dimensão
5D	Quinta dimensão
6D	Sexta dimensão
7D	Sétima dimensão

LISTA DE SÍMBOLOS

>	Maior
<	Menor
=	Igualdade
+	Adição
E	Eletronegatividade de Pauling
r_{vdW}	Raio de Van der Waals
r_{vdW}/ϵ	Razão Raio de Van der Waals por Eletronegatividade de Pauling

SUMÁRIO

PRIMEIRA PARTE	19
1 INTRODUÇÃO	20
1.1 Justificativa	21
1.2 Objetivos	21
1.2.1 Objetivo geral	21
1.2.2 Objetivos Específicos	22
2 REFERENCIAL TEÓRICO	23
2.1 Herbicidas	23
2.2 Benzamidas e Alvo Biológico	25
2.3 Relação Quantitativa Estrutura-Atividade (QSAR)	28
2.4 MIA–QSAR (<i>Multivariate Image Analysis applied to QSAR</i>)	30
2.5 Métodos de Análise Multivariada	34
2.6 Avaliação dos modelos QSAR	36
2.7 Ancoramento Molecular (<i>docking</i>)	39
3 METODOLOGIA	40
3.1 Construção dos Modelos 2, 3 e 4 usando Aug-MIA-QSAR	44
3.2 Validação dos modelos QSAR	46
3.3 Estudos de ancoramento molecular dos herbicidas benzamidicos	46
4 CONSIDERAÇÕES GERAIS	48
REFERÊNCIAS	49
SEGUNDA PARTE	54

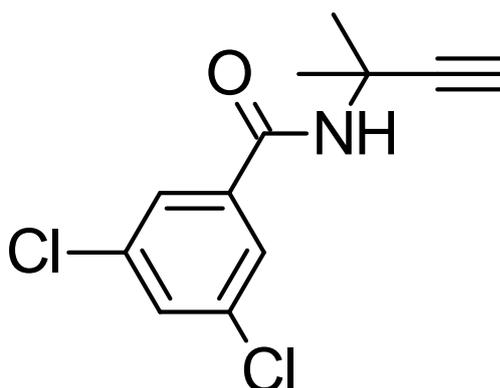
ARTIGO -----	55
ABSTRACT -----	55
1. INTRODUCTION -----	56
2. MATERIALS AND METHODS -----	57
3. RESULTS -----	60
4. DISCUSSION -----	63
5. CONCLUSIONS -----	66
REFERENCES -----	67
APPENDIX A - SUPPLEMENTARY MATERIAL -----	71

PRIMEIRA PARTE

1 INTRODUÇÃO

Na agricultura, é comum o aparecimento de plantas invasoras, o que acarreta ao agricultor o desafio de controlá-la, uma vez que elas causam um prejuízo enorme. Caso não seja controlada, a planta invasora pode dominar a cultura e causar até mesmo a morte da planta. Um recurso utilizado pelos produtores é a utilização dos herbicidas para controle ou eliminação das plantas daninhas. É comum as plantas desenvolverem resistências aos herbicidas, devido à utilização contínua desses agroquímicos. Logo, cabe aos pesquisadores sempre trabalharem no desenvolvimento de novos produtos ou no melhoramento dos herbicidas existentes. Uma classe de herbicidas presente no mercado corresponde às benzamidas; um conhecido exemplar dessa classe é a pronamida (**Figura 1**), cujo nome comercial é Kerb (Dow Agrosiences).

Figura 1 - Estrutura química do herbicida pronamida, da classe das benzamidas.



Fonte: López et al. (2015).

Sabe-se que um dos modos de ação dos herbicidas é inibir o transporte de elétrons na fotossíntese. Essa atividade está relacionada à estrutura molecular do herbicida, ou seja, modificando-se a estrutura molecular do herbicida, altera-se (aumentando ou diminuindo) a sua capacidade de inibição do fotossistema II (FSII), que é a enzima envolvida nas reações dependentes da luz, das quais depende o processo de fotossíntese.

Uma técnica utilizada para compreender a relação existente entre a estrutura química e a correspondente bioatividade (atividade biológica), com o intuito de construir um modelo matemático é o QSAR (*Quantitative Structure-Activity Relationships*) (TIMMERMAN et al., 2002). A correlação da estrutura química com a correspondente propriedade biológica ocorre

por meio de descritores moleculares, que são grandezas numéricas facilmente obtidas através de cálculos ou medidas experimentais.

A presente pesquisa possui a finalidade de encontrar uma correlação entre as estruturas moleculares de uma série de herbicidas benzamídicos e suas correspondentes capacidades inibitórias de fotossíntese, usando descritores MIA (*Multivariate Image Analysis*) (BARIGYE; FREITAS, 2016). Esses descritores moleculares correspondem aos pixels das imagens das estruturas químicas dos herbicidas (obtidas com o auxílio de programas para desenho de moléculas), cujas diferenças explicam os diferentes fatores de inibição da fotossíntese da série de compostos. As imagens são codificadas numericamente, de tal forma que os pixels apresentam valores; portanto, as imagens das estruturas químicas são compostas por pixels que variam em valor e coordenada, o que explica a variação nos valores de bioatividade. Dessa forma, pode-se compreender quais porções estruturais poderiam intensificar a inibição do FSII e, conseqüentemente, contribuir para a proposição de novos herbicidas.

1.1 Justificativa

Algumas benzamidas são conhecidas pela sua eficácia no controle de plantas daninhas. Na agricultura, a resistência das plantas aos herbicidas é comum, o que causa grande prejuízo na produção. Portanto, é imprescindível que se desenvolvam novos candidatos a herbicidas benzamídicos para o controle de pragas, uma vez que herbicidas mais eficientes podem ser utilizados em uma concentração menor, gerando, assim, menor impacto ao meio ambiente, o que pode ser iniciado por meio de métodos computacionais, tal como propõe o presente projeto.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral do projeto é compreender como a atividade de inibição fotossintética depende da estrutura química de herbicidas da família das benzamidas, bem como propor candidatos a herbicidas benzamídicos por meio da técnica de modelagem molecular MIA-QSAR.

1.2.2 Objetivos Específicos

- Construir um modelo MIA-QSAR com elevada capacidade preditiva da atividade de inibição fotossintética para uma série de herbicidas da família das benzamidas.
- Interpretar o modelo com base nos coeficientes de regressão PLS e nas projeções de variáveis importantes dos descritores MIA, com vistas a compreender quais regiões e substituintes na molécula mais afetam a bioatividade.
- Realizar estudos de *docking* molecular, para entender o modo de interação entre as moléculas do conjunto de dados e propostas e o alvo biológico;
- Divulgar os resultados por meio de publicação em revista internacional com fator de impacto relevante para a área.

2 REFERENCIAL TEÓRICO

2.1 Herbicidas

O aumento da população mundial nos últimos anos é seguido de um grande desenvolvimento tecnológico. Esse alto crescimento populacional, em junção com o consumo e a demanda por alimentos, vem exigindo dos produtores rurais a utilização de recursos como agroquímicos para garantir a produção de alimentos (JARDIM et al.,2009). O termo agrotóxico possui como sinônimos os termos praguicida, pesticida, biocida, agroquímico, defensivo agrícola, produto fitossanitário e desinfestante domissanitário (PARANÁ, 2018). Esses são definidos como produtos e agentes de processos físicos, químicos ou biológicos utilizados na produção e armazenamento e benefício da agricultura, com a finalidade de preservar as plantações de organismos vivos considerados nocivos, segundo a Lei Federal 7.802/1989 (“lei dos agrotóxicos). Resumidamente, os agroquímicos são substâncias utilizadas para proteger as plantações de organismos nocivos (LEI FEDERAL 7.802 de julho de 1989). Os agrotóxicos podem ser classificados de acordo com o tipo de praga que combatem: inseticidas (contra insetos), fungicidas (contra fungos), acaricidas (contra ácaros), herbicidas (contra plantas daninhas).

As plantas daninhas são umas das principais causas de perdas de rendimento da produção agrícola, uma vez que competem com as culturas por nutrientes, água e luz. Os herbicidas são as principais ferramentas usadas no controle de plantas daninhas na agricultura moderna. No entanto, o complexo desafio que as ervas daninhas representam, muito em razão da resistência aos herbicidas que as mesmas podem desenvolver aos agroquímicos (QUARESHY et al., 2017).

Os primeiros relatos da utilização de herbicidas surgiram por volta de 1908 e consistiam no emprego de sais de cobre e, depois, ácido sulfúrico para o controle de plantas daninhas em cereais (1993 citado por ZHINDHAL; OLIVEIRA, 2011). A partir disso, a utilização de tais produtos se intensificou, principalmente devido ao crescimento populacional. Os herbicidas são capazes de selecionar o tipo de planta em que vão agir, provocando a sua morte; o termo selecionar é utilizado no sentido da atuação destes produtos, provocando a morte da planta daninha, enquanto mantém a cultura intacta. A etimologia da palavra herbicida vem do latim *herba* (planta) e *caedere* (matar) (1993 citado por ZHINDHAL; OLIVEIRA, 2011).

Os herbicidas podem ser classificados de diversas maneiras: quanto à seletividade, translocação, época de aplicação, estrutura química, mecanismo de ação, entre outros. No presente trabalho, será dado destaque apenas quanto à estrutura química e mecanismo de ação. A classificação quanto à estrutura química é considerada limitada, uma vez que herbicidas de um mesmo grupo de compostos podem ter mecanismos de ação distintos; logo, essas duas classificações geralmente são associadas para aumentar sua utilidade (ROMAN et al., 2005). Os herbicidas podem ter vários mecanismos de ação, podendo atuar como inibidores de acetolactato sintase (ALS), do fotossistema I e II, entre diversos outros (OLIVEIRA, 2011). É importante distinguir a diferença entre mecanismo de ação e modo de ação. Mecanismo de ação é o primeiro evento metabólico (sítio de ação) que ocorre nas plantas onde o herbicida atua. Já o modo de ação é o conjunto de eventos metabólicos que resultam na expressão final do herbicida sobre as plantas, incluindo os sintomas visíveis (engloba o mecanismo de ação) (CARVALHO, 2013). Um dos mecanismos de ação dos herbicidas é a inibição do transporte de elétrons na fotossíntese, o qual resulta na remoção ou inativação de um ou mais carregadores intermediários do transporte de elétrons. As benzamidas compreendem uma classe de herbicidas que apresentam esse mecanismo de ação de inibição do fotossistema II, com o modo de ação de inibir o processo da fotossíntese (vide seção 4.2).

É importante destacar o que é necessário para que um herbicida seja eficaz e cumpra o seu papel de matar a plantas consideradas daninhas em uma cultura. De acordo com Roman et al. (2005), para ser eficiente o herbicida necessita:

- ser retido pela folhagem;
- penetrar/ultrapassar a cutícula;
- mover-se nos espaços com água ao redor da célula;
- entrar na célula, passando através da membrana celular;
- atingir o local de ação;
- ligar-se à enzima-alvo e inibi-la.

Um bom herbicida deve ainda ser seletivo, ou seja, afetar o local de ação da planta invasora, mas não na cultura, ser metabolizado ou degradado pela cultura e não ser pela planta daninha, ou seja, ele deve atingir apenas na planta alvo (ROMAN et al.,2005).

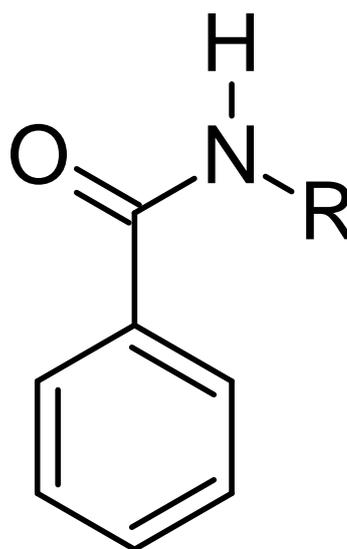
O uso excessivo de herbicidas impõe uma forte seleção para qualquer característica, permitindo que as populações de plantas sobrevivam e se reproduzam sob pressão recorrente

de herbicidas. Isso contribuiu para a evolução mundial da resistência a herbicidas em ervas daninhas (REY-CABALLERO et al.2017). A resistência a herbicidas causa perdas de safra, contaminação de sementes de ervas daninhas, valores de terra reduzidos, maiores custos de manejo mecânico e cultural de ervas daninhas e despesas adicionais com herbicidas alternativos e/ou sistemas de cultivo para manejar populações resistentes a herbicidas (NORSWORTHY et al. 2012). Logo, como o uso de herbicidas apresenta grande contribuição para a garantia de uma boa safra agrícola, o desenvolvimento e estudo de novos herbicidas deve ser constante, em razão das resistências diferentes que eventualmente apareçam.

2.2 Benzamidas e Alvo Biológico

A inibição da capacidade fotossintética tem sido considerada uma estratégia importante para a ação de herbicidas e, de fato, vários compostos que inibem a fotossíntese constituem a maior classe de herbicidas comerciais (HONDA et al.,1990). As benzamidas fazem parte de uma classe de herbicidas que interferem no processo de fotossíntese das plantas. Elas compreendem uma das classes de herbicidas utilizadas no mundo, e sua estrutura básica é mostrada na **Figura 2** (MACKAY et al., 1997).

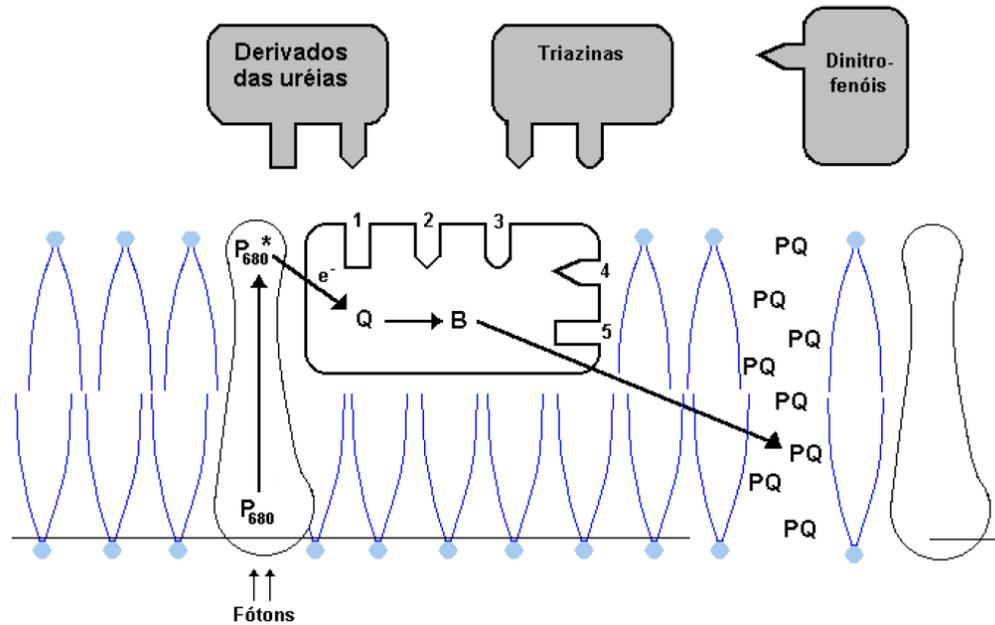
Figura 2 - Estrutura química básica de uma benzamida, precursora dos herbicidas benzamídicos.



Fonte: BENZAMIDA (2021).

ligam aos sítios 1 e 2, e herbicidas do subgrupo C3 se ligam ao sítio 4 (OLIVEIRA et al., 2011).

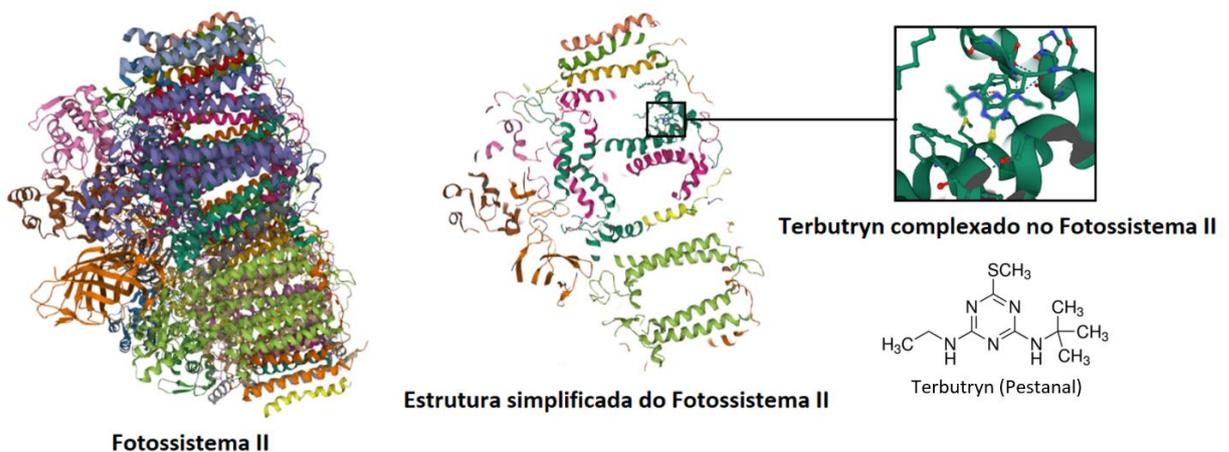
Figura 4 - Visualização dos sítios de ligação de alguns herbicidas inibidores do FS II ao complexo proteico QB na membrana dos cloroplastos



Fonte: OLIVEIRA et al.,2011

A Figura 5 mostra o herbicida terbutryn complexado ao fotossistema II (FSII) de uma cianobactéria. O terbutryn (comercialmente conhecido como Pestanal) é um herbicida triazínico, mas o seu modo e local de inibição são idênticos aos de herbicidas benzamídicos.

Figura 5 - Estrutura do fotossistema II (FSII) complexado ao herbicida terbutryn.



Fonte: PDB: 4V82 (BROSER et al.,2011).

O modo de ação completo ocorre pela inibição da fotossíntese, onde os herbicidas se ligam ao sítio de ligação da plastoquinona Q_B na proteína D1 do FSII, o qual se localiza nas membranas dos tilacóides dos cloroplastos, causando o bloqueio do transporte de elétrons de Q_A para Q_B . Resumidamente, a morte das plantas ocorre devido ao bloqueio de elétrons de Q_A para Q_B , o que interrompe a fixação de CO_2 e a produção de energia (ATP e $NADPH_2$), que são elementos essenciais para o crescimento das plantas (SILVA et al., 2013 a b).

O mecanismo de ação das benzamidas é denominado inibição do transporte de elétrons fotossintéticos (PET), o qual é quantificado através de medidas de uma reação fotossintética em membranas tilacoides, isoladas a partir de folhas ou algas unicelulares. Esse sistema de medição é considerado conveniente, confiável e preciso para avaliar a atividade herbicida de benzamidas, pois ele não sofre interferências da complexa fisiologia da planta (BARBER, 1992). O potencial inibitório é expresso em termos de pIC_{50} , isto é, o logaritmo negativo da concentração que inibe o FSII pela metade (50%).

2.3 Relação Quantitativa Estrutura-Atividade (QSAR)

A relação quantitativa estrutura-atividade - QSAR (*Quantitative Structure-Activity Relationship*) ou, ainda, o termo mais geral relação quantitativa estrutura-propriedade - QSPR (*Quantitative Structure-Property Relationship*), é resultado final de um processo que se inicia com uma descrição adequada de estruturas moleculares e termina com determinada inferência, hipótese, previsão sobre o comportamento de moléculas em sistemas ambientais, biológicos e físico-químicos em análise (TODESCHINI et al., 2020). Os estudos de QSAR é baseado no princípio de que a atividade biológica de um composto está relacionada à sua estrutura ou propriedades moleculares, e moléculas estruturalmente semelhantes podem ter atividades biológicas semelhantes (MYINT; XIE, 2010). As informações estruturais moleculares são codificadas em descritores moleculares.

Os descritores moleculares podem ser definidos como resultado de um procedimento lógico e matemático, que converte informações químicas codificadas com uma representação simbólica de uma molécula em um número útil ou resultado de algum experimento padronizado (CONSONNI; TODESCHINI, 2000).

Os descritores moleculares podem ser de dois tipos: empíricos, ou seja, as variáveis independentes do modelo são obtidas experimentalmente (descritores moleculares empíricos), ou teóricos, os quais são obtidos computacionalmente com base na estrutura molecular (descritores moleculares teóricos) (KATRITZKY et al., 2010) (ESTRADA, 2008).

Os descritores moleculares podem ser classificados quanto à sua dimensionalidade, que vai desde unidimensional (1D) até a sétima dimensão (7D), o que classifica os métodos QSAR em QSAR-1D a QSAR-7D (QUADRO 1).

Os modelos de QSAR são construídos utilizando-se a atividade biológica como variável resposta (y) (propriedade a ser predita) e os descritores moleculares como as variáveis independentes (x). A representação matemática geral para o QSAR é dada pela Equação 1:

$$y = f(x_1, x_2, \dots, x_n) \quad (1)$$

Assim, pode-se resumir que um modelo QSAR define relações matemáticas entre descritores e atividades biológicas de compostos conhecidos, para prever atividades de compostos desconhecidos (MAHALAKSHMI; JAHNAVI, 2020).

Yousefinejad e Hemmateenejad (2015) destacam que as análises QSAR apresentam pelo menos cinco etapas: (1) Seleção do grupo de moléculas a ser estudado; (2) Geração dos descritores moleculares (características estruturais) que refletem a estrutura das moléculas; (3) Seleção de bons descritores para construção do modelo de predição; (4) Construção do modelo e, por fim; (5) O modelo construído deve ser avaliado por meio dos métodos estatísticos de validação.

QUADRO 1 - Classificação das metodologias QSAR, com base na dimensionalidade do descritor.

1D-QSAR	Consiste de representações moleculares e fragmentos moleculares, os quais apresentam correlação como, por exemplo, pKa e log P.
2D-QSAR	Dados topológicos que levam em consideração o arranjo atômico em um composto, fornecendo informação sobre tamanho, forma e ramificação de uma molécula.
3D-QSAR	Inclui descritores que consideram o arranjo espacial dos átomos em uma molécula.
4D-QSAR	Considera interações entre ligante e receptor, e/ou múltiplas conformações do ligante.
5D-QSAR	Considera o efeito do receptor na conformação do ligante (encaixe induzido).
6D-QSAR	Incorpora efeitos de solvatação ao 5D-QSAR.
7D-QSAR	Considera modelos reais ou virtuais da interação ligante-receptor.

Fonte: elaborado pelo autor com base em CHERKASOV et al. (2014), DAMALE et al. (2014), CORMANICH et al. (2012), TODESCHINI et al. (2020), e POLANSKI (2009).

Em decorrência das diferentes abordagens existentes, cada qual com suas vantagens e desvantagens, é de suma importância o uso de uma técnica que consiga equilibrar a simplicidade na geração de descritores eficientes e a interpretabilidade do modelo. Nesse sentido, a metodologia denominada MIA-QSAR (Análise Multivariada de imagens aplicada em QSAR, do inglês *Multivariate Image Analysis applied to QSAR*) parece ser uma alternativa interessante para modelagem molecular (BARIGYE; FREITAS, 2016).

2.4 MIA–QSAR (*Multivariate Image Analysis applied to QSAR*)

A análise multivariada de imagens aplicada às relações quantitativas entre estrutura e atividade (*multivariate image analysis applied to quantitative structure-activity relationships* ou MIA-QSAR) é um método que se baseia no tratamento de imagens bidimensionais (2D) para modelagem molecular (FREITAS et al., 2005). Antes de se descrever especificamente sobre o uso de análise multivariada de imagens aplicado ao QSAR e entender como a metodologia MIA-QSAR funciona, é importante entender o que é uma imagem.

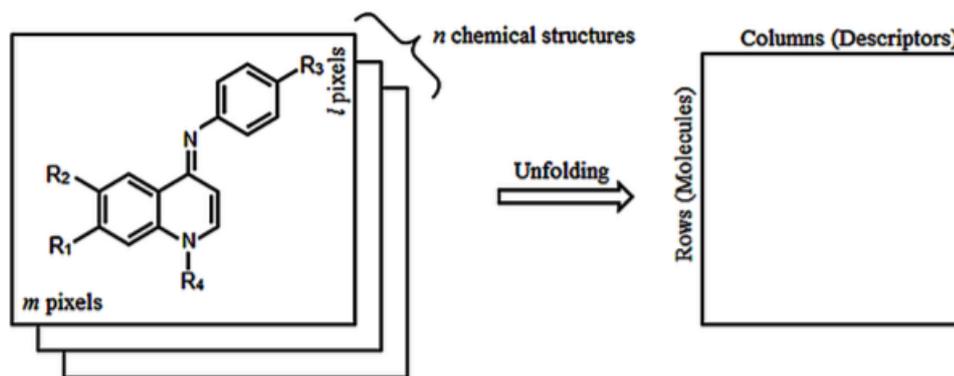
Uma imagem pode ser descrita como uma contribuição bidimensional de intensidades de luz capturadas de algum meio físico. Uma imagem possui significado visual, ou seja, representa algo. Ela é formada por pixels, os quais não assumem um valor qualquer ou não-correlacionado de intensidade em relação aos seus vizinhos, senão, dessa maneira, não se teria uma imagem. Logo, os pixels devem oferecer determinada autocorrelação espacial, para que, juntos, correspondam à imagem (GELADI et al., 1992).

Existem dois tipos de imagens: as univariadas e as multivariadas. As primeiras são compostas por duas dimensões geométricas (x e y) e estão em escala cinza; então, em uma imagem univariada, cada pixel possui uma única coordenada, para a qual é atribuída um valor de intensidade. O outro tipo de imagem corresponde às multivariadas, que representam um conjunto de imagens univariadas, que podem ser visualizadas como um arranjo de pixels, na qual cada valor de um pixel não corresponde a um valor único de intensidade de luz, mas sim à junção de diferentes bandas (canais). O sistema de cores RGB é um exemplo disso, em que cada pixel tem a contribuição de três bandas, que correspondem às cores vermelho (Red), verde (Green) e azul (Blue) (ESBENSEN; GELADI, 1989). Nesse sistema de cores RGB, o

espectro de cores é considerado um produto da contribuição de três cores. Cada cor no sistema RGB-8 bits é representada por 1 byte de memória (8 bits), o que permite 255 valores diferentes. Como um único pixel tem a contribuição das três cores, então, ele poderá assumir valores que vão de 0 (ausência de cor - preto) a 765 (contribuição das 3 cores - branco) (NUNES; FREITAS, 2013).

A estratégia MIA-QSAR foi motivada a partir do raciocínio de que, para uma dada série de estruturas químicas congruentes, a variação nas propriedades observadas é em função das subestruturas/grupos não congruentes. Essa metodologia é baseada no tratamento de imagens (estruturas químicas 2D), onde as variáveis independentes (descritores) são pixels (binários, originalmente) carregados dessas imagens. O simples alinhamento dessas estruturas resulta em um arranjo tridimensional, em razão da sobreposição das imagens 2D, mantendo-se as subestruturas coincidentes das moléculas congruentes. Em uma segunda etapa (**Figura 6**), o arranjo tridimensional é desdobrado para fornecer uma matriz de dados, onde as linhas representam os compostos e as colunas as variáveis, adequada para aplicar um método de regressão multivariada conveniente para correlacionar os descritores com o bloco y (o vetor da coluna de variáveis dependentes) (CORMANICH; FREITAS; RITTNER, 2011).

Figura 6 - Esquema utilizado para o tratamento das imagens e geração dos descritores MIA-QSAR.



Fonte: BARIGYE; FREITAS, 2016.

O primeiro modelo MIA-QSAR, atualmente denominado MIA-QSAR tradicional ou MIA-QSARt, utilizava imagens de estruturas químicas em preto e branco (conforme a **Figura 6a**) como fonte de descritores moleculares. A matriz binária de pixels era composta por valores 0 (preto) e 765 (branco), exclusivamente, onde 765 representa os espaços em branco

da imagem, enquanto o valor 0 indica os *wireframes* que compõem as estruturas químicas (FREITAS et al., 2005).

Tal metodologia, apesar de ter alcançado resultados promissores em várias análises, possui algumas limitações em função do alinhamento impreciso dos *wireframes*, o que gera variação nos dados, e também devido à representação imperfeita do tamanho e forma dos átomos e dos grupos químicos, o que fazia com que esse método falhasse em diferenciar heteroátomos e/ou tipos de átomos (NUNES; FREITAS, 2013).

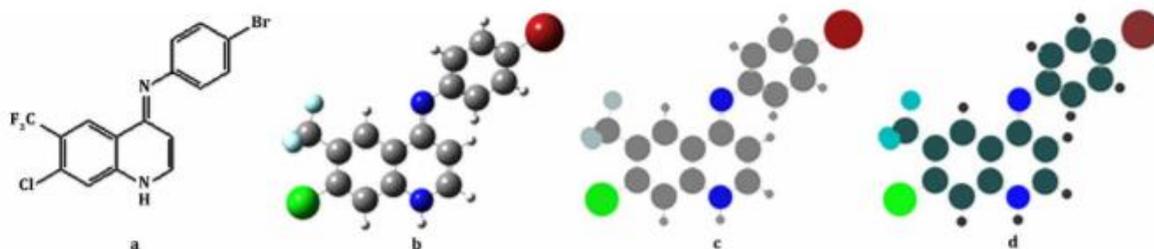
Com a finalidade de eliminar os problemas mencionados, evoluiu-se para o então denominado *aug-MIA-QSAR* (*augmented MIA-QSAR*) (NUNES; FREITAS, 2013). Essa abordagem foi cunhada através da introdução de esquemas de cores ao método tradicional MIA-QSAR, marcando uma importante etapa para a modelagem baseada em análises de imagens multivariadas. Conseqüentemente, isso possibilitou a integração de informações químicas vitais e permitiu a discriminação dos heteroátomos de forma mais significativa, além de proporcionar uma maior capacidade de modelagem (NUNES; FREITAS, 2013).

No *aug-MIA-QSAR*, os átomos são representados por círculos (correspondentes aos raios de van der Waals) de cores sólidas (oriundas do sistema RGB) para representação dos átomos. O esquema *aug-MIA-QSAR* compreende duas abordagens: o *aug-MIA-QSARvol* e *aug-MIA-QSARcolor*. Na abordagem *aug-MIA-QSARvol*, os átomos que constituem estruturas químicas são atribuídos a cores padrão (conforme estabelecido pelo programa de desenho de estrutura química, por exemplo, GaussView), com o objetivo de apenas discriminar os tipos de átomos, enquanto os raios dos círculos que representam os átomos são definidos para serem proporcionais ao raio de Van der Waals (volume). Já na abordagem *aug-MIA-QSARcolor*, as cores dos átomos são cuidadosamente escolhidas para garantir a correspondência direta às diferentes propriedades atômicas, por exemplo, eletronegatividade, raio de Van der Waals, razão raio/eletronegatividade, etc. (BARIGYE; FREITAS, 2016).

O uso de cores e volumes atômicos para diferenciar átomos e grupos químicos fez com que o método MIA-QSAR melhorasse sua habilidade de predição se comparado à metodologia convencional (BARIGYE; FREITAS, 2016). Para um entendimento mais claro, a Figura 7 expressa claramente como eram feitas as representações dos compostos segundo a primeira versão do modelo MIA-QSAR (a), de acordo com o modelo *aug-MIA-QSARvol* (b), de acordo com o modelo *aug-MIA-QSARvol* (átomos com cores sólidas) (c), e *aug-MIA-QSARcolor* (d).

Figura 7 - 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).

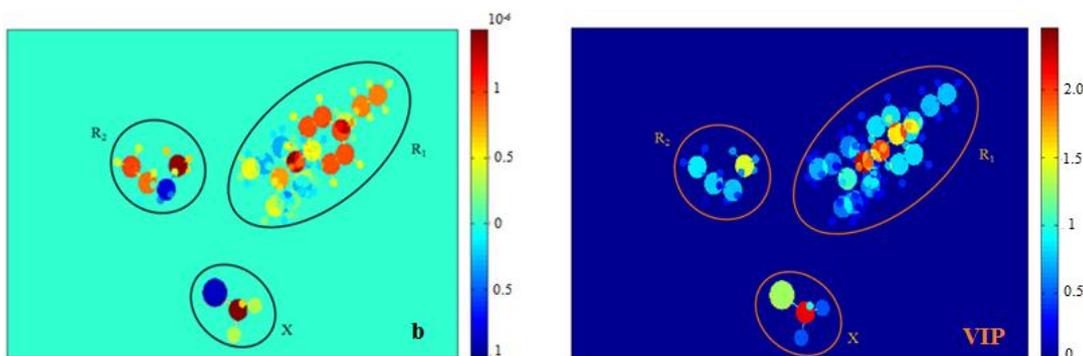


Fonte: BARIGYE; FREITAS, 2016

Novos melhoramentos no método MIA-QSAR aconteceram, com o desenvolvimento de uma ferramenta gráfica, denominada de MIA-Plot, que contribui para a interpretação química dos modelos gerados, permitindo o entendimento de como os diferentes substituintes e suas respectivas posições afetam a propriedade de interesse (BARIGYE et al., 2016). Essa nova ferramenta consiste da utilização de dois gráficos de cores, em que o primeiro é chamado de gráfico de coeficientes de regressão do PLS (**b**), e o segundo é um gráfico de *scores* da importância das variáveis em projeção (VIP) (BARIGYE et al., 2016).

O mapa de *scores* VIP descreve o quão importante um determinado descritor é para o modelo, enquanto o gráfico de coeficientes de regressão PLS (**b**) indica como um determinado descritor molecular afeta a propriedade de interesse (aumentando ou diminuindo, por exemplo). Portanto, uma análise conjunta dos gráficos VIP e **b** é recomendável (BARIGYE et al., 2016). A **Figura 8** ilustra um exemplo de utilização dos MIA-Plots.

Figura 8 - MIA-Plots baseados nos parâmetros **b** e VIP para derivados de triazinas.



Fonte: DARÉ et al., 2017

Dependendo do dado com que se está lidando, há diversas formas para a obtenção dos descritores de imagens. Quando o caso é a técnica MIA-QSAR, somente os dados relacionados com mudanças estruturais que afetam a propriedade interessada estudada são relevantes. Portanto, os descritores que correspondem aos espaços em branco comuns das imagens, bem como às subestruturas coincidentes das moléculas, são excluídos da matriz de dados. Em seguida, após a obtenção dos descritores (geralmente, centenas ou milhares), utiliza-se um método de análise multivariada adequado para regressão dos dados e, enfim, para construção do modelo de previsão.

2.5 Métodos de Análise Multivariada

A utilização de computadores para a análise das informações químicas aumentou enormemente nos últimos anos, graças aos avanços tecnológicos na área de informática. A análise de dados, principalmente no ramo da química analítica (área da química que trabalha com identificação ou quantificação de espécies ou elementos químicos), atingiu um sofisticado nível com a interface de instrumentos aos computadores, produzindo, assim, enorme quantidade de informações complexas e variadas. Um exemplo comum é o tratamento de dados das intensidades de absorção em muitos comprimentos de onda obtidas de um único espectro. Considerando essa imensa quantidade de dados, é necessária a utilização de ferramentas quimiométricas para a sua análise. A quimiometria emprega métodos matemáticos estatísticos para que, assim, consiga obter o máximo de informação química dos dados (FERREIRA et al, 1999).

A medida quantitativa de uma propriedade de interesse é uma tarefa muito comum em química e, em muitos casos, a propriedade está relacionada com a concentração de uma determinada espécie presente em um meio. Quando se utilizam técnicas instrumentais de análise, a concentração (propriedade de interesse), por exemplo, é estipulada de modo indireto por meio da relação com outra grandeza, como a absorção, em uma prática conhecida como calibração. A calibração é um processo definido como uma série de operações que relacionam medidas instrumentais com os valores correspondentes à propriedade estudada. Existem os métodos de regressão univariada, segundo os quais apenas uma medida instrumental para cada amostra é avaliada e, nos casos em que existem duas ou mais respostas instrumentais que relacionam com a propriedade de interesse, utilizam-se os métodos de calibração multivariada (BRAGA; POPPI, 2004).

Em uma análise QSAR, existe uma grande quantidade de variáveis independentes (descritores) que descrevem o sistema e, portanto, há a necessidade do uso de métodos de calibração multivariada para se obter uma equação que descreva a correlação entre os descritores e a propriedade.

Os métodos lineares de regressão multivariada relacionam duas ou mais variáveis independentes (x) à variável dependente (y) como demonstrado na Equação 2.

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

Em que b_i são os coeficientes de regressão, n o número de variáveis independentes e “ e ” o valor residual (FERREIRA, 2002; GELADI; KOWALSKI, 1986).

Os métodos de análise multivariada mais comumente empregados em QSAR consistem de aproximações qualitativas, tais como a Análise dos Componentes Principais (PCA), e quantitativas, tais como Regressão Linear Múltipla (MLR) e Mínimos Quadrados Parciais (PLS). A escolha do método mais adequado depende de diversos fatores, tais como o número de descritores envolvidos, se há ou não colinearidade (redundância) entre os descritores, se a relação existente entre eles é bem conhecida, dentre outros fatores (BEEBE et al., 1998).

No caso do presente trabalho, em razão da imensa quantidade de descritores gerados pela metodologia MIA-QSAR (*Multivariate Image Analysis applied to QSAR*) empregada e da elevada colinearidade existente entre os mesmos, a ferramenta de análise multivariada escolhida foi o PLS. Segundo Beebe et al. (1998), PLS é um método para construir modelos de previsão quando existem muitos fatores (descritores) e os mesmos possuem uma alta colinearidade. Além disso, essa ferramenta quimiométrica deve ser empregada quando o objetivo principal de prever a variável resposta, e não em tentar explicar a relação existente entre os fatores. A regressão por PLS se baseia na ideia de que, dentre uma grande quantidade de descritores, apenas alguns realmente contribuem para a variação na variável resposta (BEEBE et al., 1998). Esses descritores são conhecidos como variáveis latentes (LV) e o método de PLS tem por objetivo extraí-las e construir o modelo de predição a partir das mesmas.

De forma geral, o método PLS pode ser descrito como um método de extração das variáveis latentes da matriz \mathbf{X} . A matriz que possui as LV das amostras chama-se T e a que possui as variáveis com as informações da resposta buscada é chamada de U . As variáveis latentes são, então, utilizadas para construção das predições das respostas. T e U são

escolhidos de forma que a correlação entre os pares sucessivos dos pontos seja a maior possível. Inicialmente, a técnica PLS procura direções no espaço de variáveis que estão associadas com as maiores variações na variável resposta (TOBIAS, 2016).

Vale destacar que a escolha do número de variáveis latentes é de suma importância para construção do modelo. O número de LV geralmente é escolhido com base no decaimento do RMSE (erro quadrático médio, do inglês Root Mean Square Error) na validação cruzada do tipo *Leave-one-out*.

É importante destacar que os modelos gerados a partir desses métodos devem ser cuidadosamente validados para garantir a confiabilidade dos dados de resposta obtidos.

2.6 Avaliação dos modelos QSAR

O processo de validação é crucial nos estudos de QSAR. Esse processo busca garantir a confiabilidade dos modelos QSAR desenvolvidos, além de permitir outros tipos de análise, tais como a avaliação da qualidade dos dados de entrada, diversidade do conjunto de dados, previsibilidade em um conjunto externo, domínio de aplicabilidade e interpretabilidade mecanística (ROY et al., 2012).

Dentre os parâmetros de validação utilizados para a avaliação do modelo QSAR, 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), $r^2_{y\text{-rand}}$ (coeficiente de determinação obtido após a randomização do bloco das variáveis dependentes) e RMSE (erro quadrático médio).

Para que se possa realizar a validação de um modelo, o conjunto de dados é dividido em um subconjunto de treinamento (usado para construção do modelo) e um subconjunto teste (empregado para testar o poder preditivo do modelo). Essa divisão pode ser de forma randômica ou seguindo outro critério estatístico. Um método estatístico de amostragem muito utilizado é o Kennard-Stone. Esse algoritmo seleciona as amostras com base na maior distância euclidiana de seus *scores* em um componente principal. Essas amostras são removidas do conjunto de dados e inseridas em um subgrupo (grupo teste) (KENNARD; STONE, 1969).

As validações internas são realizadas no conjunto de treinamento e as validações externas são aplicadas no conjunto teste (ROY et al., 2012). Um tipo de validação interna muito popular é a validação *leave-one-out* (LOOCV). Nesse procedimento, uma amostra de cada vez é removida do conjunto de treinamento, a calibração é realizada utilizando as

amostras restantes e, por fim, os parâmetros de regressão obtidos são utilizados para prever a propriedade da amostra deixada de fora. As diferenças entre os valores experimentais e preditos são usadas para calcular o erro quadrático médio da validação cruzada ($RMSE_{CV}$) e o seu coeficiente de determinação (q^2) (ROY et al., 2012).

As estatísticas para avaliar o modelo de calibração consistem do erro quadrático médio da calibração ($RMSE_C$) e do coeficiente de determinação (r^2), obtidos para o conjunto de treinamento. Uma vez que a LOOCV representa certas perturbações ao modelo, é natural considerar que $r^2 > q^2$ e $RMSE_C < RMSE_{CV}$. As estatísticas mínimas aceitáveis para modelos de regressão em QSAR incluem as condições $q^2 > 0,5$ e $r^2 > 0,6$ (KIRALJ; FERREIRA, 2009).

Uma outra importante validação realizada no grupo de treinamento é nomeada de *y-randomization* (teste de randomização do bloco Y), que consiste em várias execuções (10-1000) para as quais a matriz de descritores originais \mathbf{X} , é mantida fixa, e apenas o vetor y é randomizado. Os modelos obtidos nessas condições devem ser de baixa qualidade e sem significado real, uma vez que um bom modelo obtido nessas condições indica que a correlação foi obtida ao acaso ou que o modelo está superajustado (*overfitting*) (KIRALJ; FERREIRA, 2009).

A validação externa (ou validação do conjunto teste) é considerada como uma prova de confiabilidade do modelo QSAR ou QSPR construído. Para realizar a validação externa, um conjunto denominado teste (conjunto que não foi utilizado para o desenvolvimento do modelo) é empregado para previsão e assim verificar a confiabilidade do modelo desenvolvido (ROY et al., 2012).

De acordo com a literatura, a validação externa de qualquer modelo é feita principalmente em um conjunto de teste selecionado do conjunto de dados original e a qualidade do modelo é avaliada com base em diferentes métricas de validação externa calculada a partir de um conjunto de teste (ROY et al., 2012).

Os valores de r^2_{test} (recomendáveis acima de 0,5) e $RMSE_p$ não são suficientes para atestar a validação externa de um modelo QSAR. Um critério semelhante ao $RMSE_p$ para medição absoluta do erro é o MAE (erro médio absoluto) (CHIRICO; GRAMATICA, 2011). Contudo, para alcançar uma indubitável confiança em um modelo QSAR, uma vigorosa validação externa é necessária, a qual se constitui das análises dos seguintes parâmetros: Q^2_{F1} , Q^2_{F2} , Q^2_{F3} e CCC (coeficiente de correlação de concordância), k e k' (CHIRICO; GRAMATICA, 2012).

O coeficiente de correlação de concordância (CCC) é utilizado para medir a concordância entre os dados experimentais e preditos, servindo como uma estimativa da precisão (a qual distância as observações estão da linha de ajuste) e da exatidão (quanto a linha de regressão se desvia da reta ideal, que possui inclinação 1 e que passa pela origem, ou seja, a linha de concordância). Consequentemente, qualquer divergência entre a linha de regressão e a linha de concordância resulta em um valor de CCC menor que 1, sendo que valores de $CCC > 0,85$ são considerados aceitáveis (CHIRICO; GRAMATICA, 2011).

Os parâmetros k e k' estão relacionados com a inclinação da reta. Quando as respostas observadas são plotadas no eixo y , as respostas previstas são plotadas no eixo x e a linha de regressão de melhor ajuste é desenhada definindo o intercepto em zero. Essa inclinação da linha de regressão é denominada k . O valor de k' é obtido da mesma forma, mas trocando os eixos. Os valores de k ou k' devem ser próximos de 1 (variando de 0,85 a 1,15) (CHIRICO; GRAMATICA, 2012).

Já os parâmetros Q^2_{F1} , Q^2_{F2} e Q^2_{F3} (Q^2_{Fn}), avaliam o desempenho preditivo do modelo em um conjunto externo. As previsões externas para Q^2_{Fn} são consideradas boas quando o limite usado apresenta um valor maior que 0,6 (CHIRICO; GRAMATICA, 2011).

O método de validação denominado *bootstrapping* consiste em dividir aleatoriamente os conjuntos de dados em subconjuntos de treinamentos e teste várias vezes (recomendado de 10 a 25 ciclos), obtendo assim valores médios dos parâmetros estatísticos (KIRALJ; FERREIRA, 2009). Esse procedimento avalia a robustez e estabilidade do modelo matemático construído.

Também é de grande importância para a construção de um modelo QSAR, verificar o domínio de aplicabilidade (DA) das amostras. O domínio de aplicabilidade ou domínio de aplicação consiste em uma região teórica no espaço químico que engloba tanto os descritores do modelo quanto a resposta modelada, o que permite estimar incerteza na previsão de um composto particular com base em quão semelhante é aos compostos de treinamento empregados no desenvolvimento do modelo (ROY et al., 2015). O DA é uma ferramenta importante a se considerar em qualquer modelo QSAR. O DA é muito importante para detectar amostras atípicas ou anômalas (outliers), que são os compostos que apresentam comportamento inesperado (ALVES et al., 2018).

2.7 Ancoramento Molecular (*Docking*)

É comum, juntamente com a metodologia QSAR, a utilização da técnica de ancoramento molecular, também conhecido como *docking*, o qual complementa o processo de desenvolvimento de moléculas bioativas. Por meio de cálculos de *docking*, é possível prever o modo de interação da molécula ligante na região de ligação de um alvo molecular, por exemplo, o sítio ativo da enzima específica, que pode ser inibido ou ativado. (MORGON; COUTINHO, 2007)

A classe chamada *docking* proteína-ligante é uma metodologia envolvida no planejamento de novos candidatos com possível atividade biológica. Esta metodologia possui o objetivo de quantificar o modelo de interação da molécula ligante com o seu respectivo receptor para verificar potencialidade ou inibição da atividade biológica. Isso é realizado pela otimização da estrutura do ligante no sítio de ligação do receptor, e a melhor posição recebe uma pontuação que indica a probabilidade de interação ligante-receptor (CAVASOTTO et al., 2018).

Os estudos de *docking* proteína-ligante incluem, resumidamente, as seguintes etapas: 1) a busca do receptor em um banco de dados; 2) os estudos computacionais em busca das possíveis conformações e orientações do ligante no receptor; 3) a estimativa das afinidades proteína-ligante nas posições geradas, de modo a classificá-las e identificá-las como o modo mais favorável de interação (DU et al., 2016).

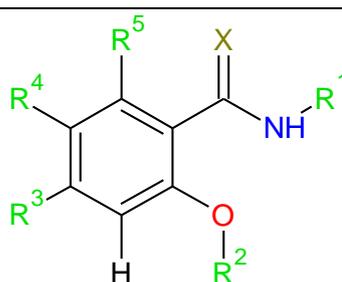
3 METODOLOGIA

O primeiro passo para a construção de um modelo QSAR consiste na escolha de um conjunto de dados, formado por uma série de compostos congêneres contendo os dados de variáveis dependentes (no caso, pIC_{50} , o logaritmo negativo da concentração que inibe o PSII - Photosystem II - pela metade), já obtidos experimentalmente e disponíveis na literatura. Algumas séries de derivados benzamídicos e seus respectivos valores de pIC_{50} encontram-se disponíveis para consulta (HONDA et al., 1990; IMRAMOVSKY et al., 2011; SHARMA, 2015) e foram utilizados para a construção dos modelos MIA-QSAR.

Ao todo, o conjunto de dados é constituído por 86 moléculas, sendo 59 compostos derivados de 3-nitro-2,4,6-trihidroxibenzamida (HONDA et al., 1990; SHARMA, 2015) e 27 compostos do grupo alquicarbamatos salicilanilidas (IMRAMOVSKY et al., 2011), totalizando 86 moléculas. Desses compostos, 80 mostraram valores de pIC_{50} precisos e foram, então, usados na modelagem MIA-QSAR (**Tabela 1**).

Foram propostas 18 novas moléculas de possíveis candidatos a herbicidas. O critério utilizado para essa proposição considerou misturar as subestruturas de moléculas que apresentavam pIC_{50} altos e que resultasse em novas estruturas químicas, ainda inexistentes no conjunto de dados (P1- P18 na tabela 1).

Tabela 1 - Compostos derivados de 3-nitro-2,4,6-trihidroxibenzamida, grupo alquicarbamatos salicilanilida e correspondentes valores de atividade inibitória do FSII (IC_{50} em mol L⁻¹).



Compound	X	R ¹	R ²	R ³	R ⁴	R ⁵	pIC_{50}
1	O	Methyl	H	OH	NO ₂	OH	5,50
2	O	Ethyl	H	OH	NO ₂	OH	6,10
3	O	Propyl	H	OH	NO ₂	OH	6,00
4	O	Butyl	H	OH	NO ₂	OH	6,10
5	O	Pentyl	H	OH	NO ₂	OH	7,00
6	O	Hexyl	H	OH	NO ₂	OH	7,20

7	O	Ciclohexyl	H	OH	NO ₂	OH	6,20
8	O	Heptyl	H	OH	NO ₂	OH	7,80
9	O	Octyl	H	OH	NO ₂	OH	8,10
10	O	Nonyl	H	OH	NO ₂	OH	8,30
11	O	Decyl	H	OH	NO ₂	OH	8,40
12	O	Undecyl	H	OH	NO ₂	OH	8,40
13	O	Tridecyl	H	OH	NO ₂	OH	8,10
14	O	Pentadecyl	H	OH	NO ₂	OH	7,30
15	O	Octadecyl	H	OH	NO ₂	OH	7,10
16	O	Benzyl	H	OH	NO ₂	OH	5,50
17	O	2-Phenetyl	H	OH	NO ₂	OH	5,30
18	O	(<i>R</i>)-1-Phenetyl	H	OH	NO ₂	OH	6,00
19	O	(<i>S</i>)-1-Phenetyl	H	OH	NO ₂	OH	6,30
20	O	3-Phenylpropyl	H	OH	NO ₂	OH	6,80
21	O	4-Phenylbutyl	H	OH	NO ₂	OH	7,40
22	O	2-Phenoxyethyl	H	OH	NO ₂	OH	6,10
23	O	2-(4-Cl-Phenoxy) ethyl	H	OH	NO ₂	OH	7,10
24	O	Phenyl	H	OH	NO ₂	OH	6,30
25	O	<i>N</i> -Me-Phenyl	H	OH	NO ₂	OH	^a
26	O	2-Cl-Phenyl	H	OH	NO ₂	OH	5,90
27	O	3-Cl-Phenyl	H	OH	NO ₂	OH	6,70
28	O	4-Cl-Phenyl	H	OH	NO ₂	OH	7,10
29	O	4-Br-Phenyl	H	OH	NO ₂	OH	7,20
30	O	2-F-Phenyl	H	OH	NO ₂	OH	6,00
31	O	3-F-Phenyl	H	OH	NO ₂	OH	6,60
32	O	4-F-Phenyl	H	OH	NO ₂	OH	6,40
33	O	3,5-Cl ₂ -Phenyl	H	OH	NO ₂	OH	6,70
34	O	3,4-Cl ₂ -Phenyl	H	OH	NO ₂	OH	7,20
35	O	2,4-Cl ₂ -Phenyl	H	OH	NO ₂	OH	7,00
36	O	2,3-Cl ₂ -Phenyl	H	OH	NO ₂	OH	5,90
37	O	2-CF ₃ -Phenyl	H	OH	NO ₂	OH	^a
38	O	3-CF ₃ -Phenyl	H	OH	NO ₂	OH	6,90
39	O	4-CF ₃ -Phenyl	H	OH	NO ₂	OH	7,60
40	O	2-Me-Phenyl	H	OH	NO ₂	OH	5,80
41	O	3-Me-Phenyl	H	OH	NO ₂	OH	6,60
42	O	4-Me-Phenyl	H	OH	NO ₂	OH	6,10

43	O	4-Et-Phenyl	H	OH	NO ₂	OH	6,50
44	O	4- <i>i</i> Pr-Phenyl	H	OH	NO ₂	OH	6,70
45	O	4-Pr-Phenyl	H	OH	NO ₂	OH	7,00
46	O	2-MeO-Phenyl	H	OH	NO ₂	OH	6,20
47	O	3-MeO-Phenyl	H	OH	NO ₂	OH	6,40
48	O	4-MeO-Phenyl	H	OH	NO ₂	OH	5,90
49	O	3,4-(MeO) ₂ -Phenyl	H	OH	NO ₂	OH	^a
50	O	4-NMe ₂ -Phenyl	H	OH	NO ₂	OH	^a
51	O	4-NO ₂ -Phenyl	H	OH	NO ₂	OH	6,70
52	S	Ethyl	H	OH	NO ₂	OH	6,60
53	S	Butyl	H	OH	NO ₂	OH	7,00
54	S	Hexyl	H	OH	NO ₂	OH	7,90
55	S	Heptyl	H	OH	NO ₂	OH	8,40
56	S	Octyl	H	OH	NO ₂	OH	8,70
57	S	Nonyl	H	OH	NO ₂	OH	8,50
58	S	Decyl	H	OH	NO ₂	OH	8,30
59	S	Phenyl	H	OH	NO ₂	OH	7,00
60	O	3-Cl-Phenyl	Ethyl	H	Cl	H	4,81
61	O	3-Cl-Phenyl	Butyl	H	Cl	H	4,87
62	O	3-Cl-Phenyl	Pentyl	H	Cl	H	5,27
63	O	3-Cl-Phenyl	Hexyl	H	Cl	H	5,27
64	O	3-Cl-Phenyl	Heptyl	H	Cl	H	5,31
65	O	3-Cl-Phenyl	Octyl	H	Cl	H	5,00
66	O	3-Cl-Phenyl	Nonyl	H	Cl	H	4,81
67	O	3-Cl-Phenyl	Decyl	H	Cl	H	4,80
68	O	3-Cl-Phenyl	Undecyl	H	Cl	H	4,74
69	O	4-Cl-Phenyl	Ethyl	H	Cl	H	5,01
70	O	4-Cl-Phenyl	Butyl	H	Cl	H	5,02
71	O	4-Cl-Phenyl	Pentyl	H	Cl	H	5,20
72	O	4-Cl-Phenyl	Hexyl	H	Cl	H	5,20
73	O	4-Cl-Phenyl	Heptyl	H	Cl	H	5,15
74	O	4-Cl-Phenyl	Octyl	H	Cl	H	5,13
75	O	4-Cl-Phenyl	Nonyl	H	Cl	H	5,10
76	O	4-Cl-Phenyl	Decyl	H	Cl	H	5,08
77	O	4-Cl-Phenyl	Undecyl	H	Cl	H	4,98
78	O	3,4-Cl-Phenyl	Ethyl	H	Cl	H	4,08

79	O	3,4-Cl-Phenyl	Butyl	H	Cl	H	4,25
80	O	3,4-Cl-Phenyl	Pentyl	H	Cl	H	4,00
81	O	3,4-Cl-Phenyl	Hexyl	H	Cl	H	3,84
82	O	3,4-Cl-Phenyl	Heptyl	H	Cl	H	3,77
83	O	3,4-Cl-Phenyl	Octyl	H	Cl	H	3,36
84	O	3,4-Cl-Phenyl	Nonyl	H	Cl	H	3,30
85	O	3,4-Cl-Phenyl	Decyl	H	Cl	H	^a
86	O	3,4-Cl-Phenyl	Undecyl	H	Cl	H	^a
P1	S	Undecyl	H	OH	NO ₂	OH	
P2	S	Tridecyl	H	OH	NO ₂	OH	
P3	O	Octyl	Heptyl	OH	NO ₂	OH	
P4	O	Nonyl	Heptyl	OH	NO ₂	OH	
P5	O	Decyl	Heptyl	OH	NO ₂	OH	
P6	O	Undecyl	Heptyl	OH	NO ₂	OH	
P7	S	Octyl	Heptyl	OH	NO ₂	OH	
P8	S	Nonyl	Heptyl	OH	NO ₂	OH	
P9	S	Decyl	Heptyl	OH	NO ₂	OH	
P10	S	Undecyl	Heptyl	OH	NO ₂	OH	
P11	S	Octyl	Pentyl	OH	NO ₂	OH	
P12	S	Octyl	Hexyl	OH	NO ₂	OH	
P13	S	Nonyl	Pentyl	OH	NO ₂	OH	
P14	S	Nonyl	Hexyl	OH	NO ₂	OH	
P15	S	Decyl	Pentyl	OH	NO ₂	OH	
P16	S	Decyl	Hexyl	OH	NO ₂	OH	
P17	S	Undecyl	Pentyl	OH	NO ₂	OH	
P18	S	Undecyl	Hexyl	OH	NO ₂	OH	

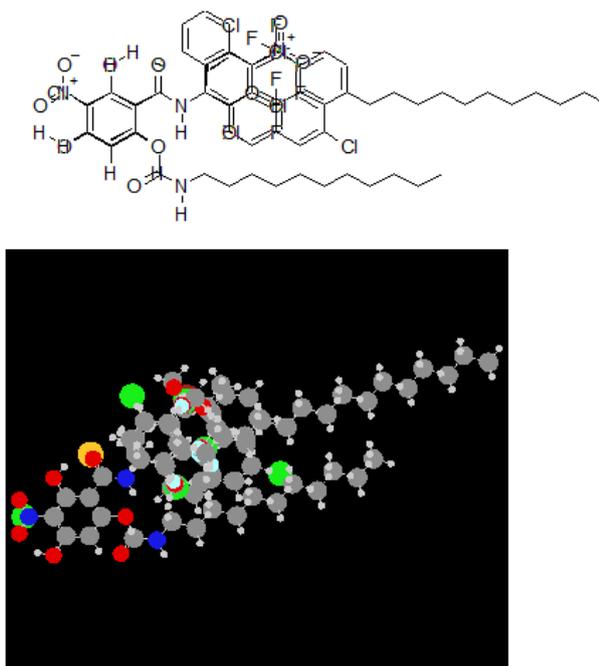
^a Inaccurate experimental values.

Fonte: (IMRAMOVSKY et al., 2011, HONDA et al., 1990; SHARMA, 2015).

O **Modelo 1** foi construído utilizando a abordagem tradicional MIA-QSAR, segundo o qual as 80 moléculas foram desenhadas usando o programa ChemSketch (ACD/ChemSketch, 2018), sendo que cada imagem foi alinhada em um espaço com dimensões predefinidas (435 × 168 pixels) no programa Microsoft Windows Paint e salvas como bitmaps (bmp). Após o alinhamento das 80 imagens correspondentes às 80 estruturas químicas (**Figura 8**), o arranjo tridimensional de dimensão 80 × 435 × 168 foi desdobrado para uma matriz de pixel **X** (80 × 73080), em que cada linha corresponde a uma molécula e as colunas correspondem aos

descritores (valores de pixel em cada coordenada das imagens originais). Posteriormente, as colunas de variância zero foram removidas para reduzir a dimensionalidade da matriz para 80×3144 . O processamento da imagem química e da matriz de dados foi realizado usando o programa Chemoface (NUNES et al., 2012). O conjunto de dados foi dividido em um subconjunto de treinamento (75% das amostras) e um de teste (25% das amostras) utilizando o algoritmo Kennard-Stone. O procedimento de calibração foi feito utilizando-se PLS, sendo que o número de variáveis latentes foi escolhido analisando-se o decaimento do RMSE (erro quadrático médio, do inglês Root Mean Square Error) na validação cruzada do tipo *leave-one-out*.

Figura 9 - Imagens sobrepostas para uso na construção do Modelo 1 (imagem superior) e dos Modelos 2-4 (imagem inferior).



Fonte: Elaborado pela Autora (2021)

3.1 Construção dos Modelos 2, 3 e 4 usando aug-MIA-QSAR

Para a construção dos **Modelos 2, 3 e 4**, as moléculas foram desenhadas no programa GaussView (Dennington et al., 2008), de forma que as moléculas ficassem perfeitamente alinhadas, para garantir que a sobreposição das subestruturas coincidentes das moléculas fosse exatamente congruente (**Figura 8**). As imagens tinham a dimensão de 354×300 pixels cada e

foram salvas como bitmaps (bmp) no Microsoft Windows Paint. O arranjo tridimensional de tamanho $80 \times 354 \times 300$ foi desdobrado em uma matriz \mathbf{X} de dimensão 80×106200 , em que cada linha corresponde a uma molécula e cada coluna a um descritor.

As colunas com variância igual a zero foram excluídas, resultando em uma matriz de dimensão 80×10631 . No GaussView, os átomos nas moléculas são representados por círculos de tamanhos proporcionais aos respectivos raios de Van der Waals e apresentam diferentes cores. Essas cores, constituídas de pixels com valores contidos na escala RGB (de 0 a 765), foram substituídas por valores quimicamente relevantes (**Tabela 2**), ou seja, valores proporcionais às grandezas atômicas como raio de Van der Waals (r_{vdw} , que deu origem ao **Modelo 2**), eletronegatividade de Pauling (ϵ , que deu origem ao **Modelo 3**) e a razão r_{vdw}/ϵ (que deu origem ao **Modelo 4**). Novamente, os grupos de treinamento (75% das moléculas) e teste (25% das moléculas) foram selecionados utilizando o algoritmo Kennard-Stone. Os modelos de calibração foram obtidos utilizando regressão PLS e a escolha do número ótimo de variáveis latentes escolhido baseou-se no decaimento do RMSE da validação cruzada *leave-one-out* (LOOCV).

Para esses três modelos foram construídas as ferramentas gráficas, denominadas MIA-plot, utilizando o programa Chemoface (NUNES et al., 2012). Os MIA-plots contribuem para a interpretação química dos modelos gerados, permitindo o entendimento de como os diferentes substituintes e suas respectivas posições afetam a propriedade de interesse (BARIGYE et al., 2016). Essa ferramenta consiste da utilização de dois gráficos de cores, em que o primeiro é chamado de gráfico de coeficientes de regressão do PLS (**b**), e o segundo é um gráfico de *scores* da importância das variáveis em projeção (VIP) (BARIGYE et al., 2016).

O mapa de *scores* VIP descreve o quão importante um determinado descritor é para o modelo, enquanto o gráfico de coeficientes de regressão PLS (**b**) indica como um determinado descritor molecular afeta a propriedade de interesse (aumentando ou diminuindo, por exemplo). Portanto, uma análise conjunta dos gráficos VIP e **b** é recomendável (BARIGYE et al., 2016).

Tabela 2 - Escala de cores RGB dos átomos obtidos no GaussView e os valores quimicamente relevantes pelos quais os valores de pixels foram substituídos.

Átomo/Ligação Química	Red	Green	Blue	Total	rvaw	ϵ	rvaw/ ϵ
Carbono	142	142	142	426	170	255	667
Hidrogênio	204	204	204	612	120	220	545
Nitrogênio	25	25	229	279	155	304	510
Oxigênio	229	0	0	229	152	344	442
Enxofre	255	198	40	493	180	258	698
Flúor	178	255	255	688	147	400	367
Cloro	25	239	25	289	175	316	554
Bromo	165	33	33	231	185	296	625
Ligação Química	205	204	204	613	0.1	0.1	0.1

Fonte: Elaborado pelo autor (2021).

3.2 Validação dos modelos QSAR

A qualidade dos modelos construídos foi avaliada pelos valores de RMSEc, r^2 , RMSE_{y-rand}, r^2_{y-rand} , r^2_p , RMSE_{cv}, q^2 , RMSE_p, r^2_{pred} , r^2_m , MAE, Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , CCC, k e k' . Adicionalmente, a robustez e estabilidade dos modelos foram verificadas por meio de *bootstrapping* (20 ciclos). O domínio de aplicabilidade do conjunto de dados (para verificar se amostras pertencem a um mesmo espaço químico) foi avaliado plotando-se o gráfico de *sample leverages* \times Student's residuals.

3.3 Estudos de ancoramento molecular dos herbicidas benzamidicos

Um estudo de *Docking* molecular foi realizado para explicar a afinidade de ligação dos herbicidas benzamidicos com a proteína FSII. A estrutura do supercomplexo *Spinacia oleracea* PSII-LHCII adquirida pelo método de microscopia eletrônica usada no estudo, com resolução de 3,2 Å, foi obtida do Protein Data Bank (código PDB: 3JCU) (WEI et al., 2016). Esta estrutura exibe uma rica coleção de detalhes atômicos: 16 clorofilas, 4 feofitinas, 8 beta-carotenos, 2 plastoquinonas, 2 protoporfirina IX contendo Fe, 2 aglomerados de Ca-Mn₄-O₅, 4 íons cloreto, 2 íons Fe (II) e 2 íons de bicarbonato. Os ligantes preparados foram encaixados no local ativo D1 do fotossistema II (psbA, EC: 1.10.3.9) usando o software Molegro Virtual

Docker (THOMSEN; CHRISTENSEN, 2006), um programa para prever a conformação de ligação mais provável de um ligante para uma macromolécula (CUNHA et al., 2011; RAMALHO et al., 2012).

A função de pontuação de energia empregada foi o MolDock Score com resolução de 0,3Å. O algoritmo de busca utilizado foi o MolDock SE. Além disso, o número de cálculos foi definido em 100, em que cada cálculo retorna uma pose (conformação).

4 CONSIDERAÇÕES GERAIS

O presente estudo consistiu de um estudo de dois conjuntos de compostos congêneros que foram reunidos em um único conjunto de 80 derivados benzamídicos, que forneceram modelos MIA-QSAR consistentes e confiáveis. Os descritores MIA conseguiram explicar propriedades topoquímicas e topoestruturais, com base em gráficos MIA que mapeiam as porções moleculares que mais afetam a atividade inibitória de FSII.

Com a mistura das subestruturas dos compostos mais promissores dentro do conjunto de dados, uma série de candidatos a herbicida de (tio)benzamida foi proposta (18 propostas). De acordo com os modelos de previsão, a melhor proposta é, provavelmente, um derivado de tiobenzamida contendo cadeia alifática C11, grupos nitro e hidroxila como substituintes. O papel desses substituintes foi racionalizado por meio de estudos de *docking*, que revelaram importantes interações ligante-enzima.

Além disso, as energias de interação intermolecular obtidas a partir dos estudos de *Docking* estão em concordância razoável com os valores de pIC_{50} previstos, fornecendo assim, uma validação adicional dos modelos MIA-QSAR.

Finalmente, essas descobertas podem ser úteis para projetar novos candidatos a agroquímicos e com desempenho herbicida melhorado, especialmente contra plantas daninhas resistentes.

REFERÊNCIAS

- ACD/CemSketch 2017.2.1, 2018. Advanced Chemistry Development, Inc., Toronto.
- ALVES, V. et al. QUIMIOINFORMÁTICA: uma introdução. **Química Nova**, [S.L.], v. 41, n. 2, p. 202-212, 2018. Sociedade Brasileira de Química (SBQ). <http://dx.doi.org/10.21577/0100-4042.20170145>.
- BARBER, J. **The photosystems: Structure, function and molecular biology**, Elsevier, Amsterdam, 1992.
- BARIGYE, S. J. et al. MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. **RSC Advances**, v. 6, n. 55, p. 49604-49612, 2016. Available at: < <http://dx.doi.org/10.1039/C6RA09593C> >.
- BARIGYE, S. J.; FREITAS, M. P. Ten Years of the MIA-QSAR Strategy: Historical Development and Applications. **International Journal of Quantitative Structure-Property Relationships**, v. 1, n. 1, p. 64–77, 2016.
- BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide**, Wiley, New York, 1998.
- BENZAMIDA. Disponível em: https://www.merckmillipore.com/BR/pt/product/Benzamide,MDA_CHEM-802191. Acesso em: 05 fev. 2021.
- BRAGA, J. W. B.; POPPI, R. J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, [S.L.], v. 27, n. 6, p. 1004-1011, dez. 2004. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-40422004000600027>.
- BRASIL. **Decreto nº 4.074, de 4 de janeiro de 2002**. Regulamenta a Lei nº 7.802, de 11 de julho de 1989, que dispõe sobre a pesquisa, a experimentação, a produção, a embalagem e rotulagem, o transporte, o armazenamento, a comercialização, a propaganda comercial, a utilização, a importação, a exportação, o destino final dos resíduos e embalagens, o registro, a classificação, o controle, a inspeção e a fiscalização de agrotóxicos, seus componentes e afins, e dá outras providências. Brasília: Diário Oficial da União, [2002]. Disponível em: <<https://bit.ly/2IGbNh7> >. Acesso em: 31 jan. 2021.
- BROSER, M. et al., Structural Basis of Cyanobacterial Photosystem II Inhibition by the Herbicide Terbutryn, **Journal of Biological Chemistry**, v. 286, n. 18, p. 15964-15972, 2011.
- CARVALHO, L. B.de. **Herbicidas**. Lages: Edição do Autor, 2013. 72 p.
- CAVASOTTO, C. et al. Computational chemistry in drug lead discovery and design. **International Journal of Quantum Chemistry**, v. 119, p. 1-19, 2018.
- CHERKASOV, A. et al QSAR Modeling: where have you been? where are you going to? **Journal of Medicinal Chemistry**, [S.L.], v. 57, n. 12, p. 4977-5010, 6 jan. 2014. American Chemical Society (ACS). <http://dx.doi.org/10.1021/jm4004285>.

CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models: How to Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. **Journal of Chemical Information and Modeling**, 51, n. 9, p. 2320-2335, 2011/09/26 2011.

CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. **Journal of Chemical Information and Modeling**, 52, n. 8, p. 2044-2058, 2012/08/27 2012.

CONSONNI, V.; TODESCHINI, R. Methods and Principles of Medicinal Chemistry. In: Mannhold, R.; Kubibiyi, H. e Timmerman, H. **Handbook of Molecular Descriptors**, Wiley-VCH publishers, 2000.

CORMANICH, R. A.; 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**, [S.L.], v. 22, n. 4, p. 637-642, 2011. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-50532011000400004>.

CORMANICH, R. A.; NUNES, C. A.; FREITAS, M. P. Desenhos de estruturas químicas correlacionam-se com propriedades biológicas: MIA-QSAR. **Química Nova**, [S.L.], v. 35, n. 6, p. 1157-1163, 2012. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-40422012000600017>.

CUNHA, E.F.F. et al., Molecular modeling of the *Toxoplasma gondii* adenosine kinase inhibitors. **Medicinal Chemistry Research**, V. 21, 590-600, 2012

DAMALE, M. et al. Recent Advances in Multidimensional QSAR (4D-6D): **A Critical Review**. Mini-Reviews in Medicinal Chemistry, v. 14, n. 1, p. 35-55, 2014.

DARÉ, J. K.; SILVA, C. F.; FREITAS, M. P. Revealing chemophoric sites in organophosphorus insecticides through the MIA-QSPR modeling of soil sorption data. **Ecotoxicology and Environmental Safety**, v. 144, p. 560-563, 2017/10/01/ 2017. ISSN 0147-6513. Available at: <
<http://www.sciencedirect.com/science/article/pii/S0147651317304074> >.

DENNINGTON, R.D, KEITH, T.A. MILLAM, M.J., *GaussView 5.0*, Wallingford, CT 2008. DU, X. et al. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. **International Journal of Molecular Sciences**, v. 17, n. 2, p. 144, 2016.

ESBENSEN, K.; GELADI, P. Strategy of multivariate image analysis (MIA). **Chemometrics and Intelligent Laboratory Systems**, v. 7, n. 1-2, p. 67-86, 1989.

ESTRADA, E. Quantum-chemical foundations of the topological substructural molecular design. **Journal of Physical Chemistry A**, v. 112, n. 23, p. 5208-5217, 2008.

FERREIRA, M.M. C. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, [S.L.], v. 22, n. 5, p. 724-731, set. 1999. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-40421999000500016>.

FERREIRA, M. Multivariate QSAR. **Journal of the Brazilian Chemical Society**, Campinas, v. 13, n. 6, p. 742-753, 2002.

FREITAS, M. et al MIA-QSAR: A simple 2D image-based approach for quantitative structure-activity relationship analysis. **Journal of Molecular Structure**, v. 738, p. 149-154, 2005.

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

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, Louisville, v. 185, p. 1-17, 1986.

HONDA, I. et al. Structure-activity relationship of 3-nitro-2,4,6-tri-hydroxybenzamide derivatives in photosynthetic electron transport inhibition, **Agricultural and Biological Chemistry**, 1990, 54, 1227-1233.

IMRAMOVSKY, A. et al Inhibiting efficiency of 4-chloro-2-(chlorophenylcarbamoyl)phenyl alkylcarbamates. **Bioorganic & Medicinal Chemistry Letters**, [S.L.], v. 21, n. 15, p. 4564-4567, ago. 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.bmcl.2011.05.118>.

JARDIM, I. C. S F.; ANDRADE, J. de A.; QUEIROZ, S. C. do N. de. Resíduos de agrotóxicos em alimentos: uma preocupação ambiental global - um enfoque às maçãs. **Química Nova**, [S.L.], v. 32, n. 4, p. 996-1012, 2009. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-40422009000400031>.

KATRITZKY, A. et al. Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. **Chemical reviews**, v. 110, n. 10, p. 5714-5789, 2010.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. *Technometrics*, v. 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**, [S.L.], v. 20, n. 4, p. 770-787, 2009. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-50532009000400021>.

LÓPEZ, B. P. et al. Agentes antimetabólicos en la obtención de plantas doble-haploides. **Reduca: Biología**, Madrid, v. 2, n. 7, p. 12-18, jan. 2015.

MACKAY, D.; SHIU, W.-Y.; MA, K.-C. **Illustrated Handbook of Physical-Chemical Properties and Environmental fate for Organic Chemicals**. Lewis Publishers, New York, 1997.

MAHALAKSHMI, P. S.; JAHNAVI, Y. A REVIEW ON QSAR STUDIES. **International Journal of Advances in Pharmacy and Biotechnology**, [S.L.], v. 6, n. 2, p. 19-23, 20 jun. 2020. MYBO Group. <http://dx.doi.org/10.38111/ijapb.20200602004>.

- MORGON, N.H.; COUTINHO, K. **Methods of theoretical chemistry and molecular modeling**. Publisher Physics Bookstore, São Paulo, 2007.
- MYINT, K.Z.; XIE, X.Q. Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods. **International Journal of Molecular Sciences**, [S.L.], v. 11, n. 10, p. 3846-3866, 2010. MDPI AG. <http://dx.doi.org/10.3390/ijms11103846>.
- NORSWORTHY, J. et al. Reducing the Risks of Herbicide Resistance: best management practices and recommendations. **Weed Science**, [S.L.], v. 60, n. 1, p. 31-62, 2012. Cambridge University Press (CUP). <http://dx.doi.org/10.1614/ws-d-11-00155.1>.
- NUNES, C. A. et al. Chemoface: A novel free user-friendly interface for chemometrics. **Journal of the Brazilian Chemical Society**, v. 23, n. 11, p. 2003–2010, 2012. <http://dx.doi.org/10.1074/jbc.M110.215970>.
- NUNES, C. A.; FREITAS, M. P. Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors. **European Journal of Medicinal Chemistry**, v. 62, p. 297–300, 2013.
- OLIVEIRA, R. S. Jr. et al. **Biologia e Manejo de Ervas Daninhas**, Omnipax, Curitiba, 2011.
- PARANÁ. Pevaspea. Governo do Estado do Paraná. **Intoxicação Agudas por Agrotóxicos: Material Técnico Intoxicações Agudas por Agrotóxicos. Atendimento Inicial do Paciente Intoxicado**. Paraná: Secretara de Saúde, 2018.
- POLANSKI, J. Receptor Dependent Multidimensional QSAR for Modeling Drug - Receptor Interactions. **Current Medicinal Chemistry**, Sharjah, v. 16, n. 25, p. 3243–3257, 2009.
- QUARESHY, M. et al. A cheminformatics review of auxins as herbicides. **Journal of Experimental Botany**, [S.L.], v. 69, n. 2, p. 265-275, 6 set. 2017. Oxford University Press (OUP). <http://dx.doi.org/10.1093/jxb/erx258>.
- RAMALHO, T.C. et al. Development of new acetylcholinesterase reactivators: Molecular modeling versus in vitro data. **Chem.-Biol. Int.**, 2012. <https://dx.doi.org/10.1016/j.cbi.2010.05.016>.
- REY-CABALLERO, J. et al. Management Options for Multiple Herbicide-Resistant Corn Poppy (*Papaver rhoeas*) in Spain. **Weed Science**, [S.L.], v. 65, n. 2, p. 295-304, mar. 2017. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/wsc.2016.38>.
- ROMAN, E. S. et al. **Como funcionam os herbicidas: da biologia à aplicação**. 21. ed. Passo Fundo: Gráfica Editora Berthier, 2005. 156 p.
- ROY, K. et al. Comparative Studies on Some Metrics for External Validation of QSPR Models. **Journal of Chemical Information and Modeling**, [S.L.], v. 52, n. 2, p. 396-408, 17 jan. 2012. American Chemical Society (ACS).
- ROY, K. et al. On a simple approach for determining applicability domain of QSAR models. **Chemometrics And Intelligent Laboratory Systems**, [S.L.], v. 145, p. 22-29, jul. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.chemolab.2015.04.013>.

SHARMA, M.C. Identification of 3-Nitro-2,4,6-trihydroxybenzamide Derivatives as Photosynthetic Electron Transport Inhibitors by QSAR and Pharmacophore Studies. **Interdisciplinary Sciences: Computational Life Sciences**, [S.L.], v. 8, n. 2, p. 109-121, 6 ago. 2015. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s12539-015-0019-9>.

SILVA, I. P. D. F. et al. Herbicidas inibidores do fotossistema II – parte I / Photosystem II inhibitor herbicides - part I. **Revista Brasileira de Engenharia de Biosistemas**, v. 7, n. 1, p. 1-11, 2013a.

SILVA, I. P. D. F. et al. Herbicidas inibidores do fotossistema II – parte II / Photosystem II inhibitor herbicides - part II. **Revista Brasileira de Engenharia de Biosistemas**, v. 7, n. 1, p. 12-22, 2013b.

THOMSEN, R.; CHRISTENSEN, M.H. MolDock: A new technique for high-accuracy molecular docking. **J. Med. Chem.** p. 3315-3321, 2006. <https://dx.doi.org/10.1021/jm051197e>.

TIMMERMAN, H. et al. **Handbook of Molecular Descriptors**, Wiley-VCH, Weinheim, 2002.

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

TODESCHINI, R. et al. 4.25 - Chemometrics for QSAR Modeling. In: BROWN, S. D.; TAULER, R., et al (Ed.). **Comprehensive Chemometrics (Second Edition)**. Oxford: Elsevier, 2020. p. 599-634.

WEI, X., SU, X., CAO, P. et al. Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. **Nature** 534, 69–74 (2016). <https://doi.org/10.1038/nature18020>

WEED S. Inhibition of photosynthesis at PSII. Disponível em: <https://www.weedscience.org/Summary/HRACPoster.aspx>. Acesso em: 10 jan. 2022.

YOUSEFINEJAD, S.; HEMMATEENEJAD, B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 177–204, 2015.

SEGUNDA PARTE

ARTIGO

Artigo submetido para publicação.

MIA-QSAR study of the structural merging of (thio)benzamide herbicides with photosynthetic system II inhibitory activities

Ingrid V. Pereira,^a Joyce K. Daré,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

^a Departamento de Química, Instituto de Ciências Naturais, Universidade Federal de Lavras, 37200-900, Lavras, MG, Brazil. E-mail: matheus@ufla.br

Word count: 4446

ABSTRACT

Benzamide herbicides consist of a class of photosynthetic system II (PSII) inhibitors widely used for weed control. However, the development of resistance by these weeds to the known herbicides requires an ongoing search for new agrochemicals. We report the combination of two congeneric series of (thio)benzamide herbicides into a single data set and subsequent modeling of their herbicidal activities against PSII using MIA-QSAR. The robust and predictive models were used to estimate the pIC₅₀ of new agrochemical candidates, which were proposed based on a chemical mixing of the substructures of the most active compounds present in the data set. The chemical features affecting the herbicidal activities were analyzed using MIA contour maps, whereas the ligand-enzyme interactions responsible for the binding affinities were rationalized through docking studies. The proposed compound possessing a thiobenzamide moiety and C-11 chain, H, NO₂, OH, and OH as variable substituents was the most promising alternative.

Keywords: Structure-activity; (thio)benzamides; herbicidal activity; photosystem II.

1. INTRODUCTION

The use of herbicides in agriculture for weed control has an enormous economic impact, *e.g.* its net global economic benefits at the farm level amounted to \$18.9 billion in 2018 and \$255.1 billion for the period 1996-2018 (Brookes and Barfoot, 2020). However, the continuing development of new agrochemicals as a diversification strategy for crop management is necessary due to the ongoing occurrence of herbicide-resistant plants (Storkey et al., 2019). Amides correspond to an important class of herbicides, such as the benzamide Pronamide (Figure 1), which is widely used for both preemergence and early postemergence control of grasses, and preemergence control of certain broadleaf weeds.

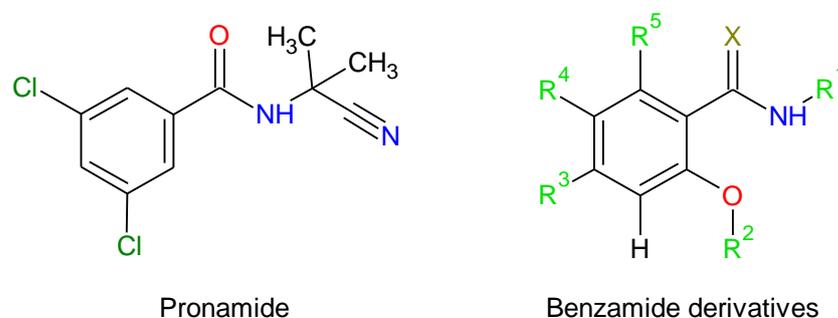


Figure 1. Commercial Pronamide herbicide and the benzamide derivatives studied herein. Changes in X (O or S) and R explain the variance in the herbicidal activity of this class of compounds.

Inhibition of photosynthetic electron transport is a key mechanism for herbicidal activity. Benzamide herbicides are within a class of compounds that block this transport by binding to the D1 subunit of the photosystem II (PSII) complex, thus inhibiting the reduction of plastoquinone at the secondary quinone acceptor site (Q_B -site) and its reoxidation in subsequent steps. The generation of reactive oxygen species due to the plastoquinone reduction blocking may cause extensive damage in PSII and eventually kills the plant (Oettmeier, 2003). Acetanilide derivatives are also well-known PS-II inhibitors of $C_2^{(7)}$ group (Devine et al., 1993). Computational approaches such as quantitative structure-activity relationships (QSAR) are inexpensive tools for understanding how changes in the chemical structure of an herbicide influence the respective biological response. This correlation may be achieved using molecular descriptors capable of encoding the interaction mode of a ligand with the binding site.

This study reports the QSAR modeling of the herbicidal activities of two series of (thio)benzamide derivatives grouped into a single data set (Figure 1) (Honda et al.,1990; Imramovsky et al., 2011; Sharma, 2016). The herbicidal activity was described in terms of the negative logarithm of the concentration of the (thio)benzamide derivative to inhibit 50% of the photosynthetic electron transport (PET) in chloroplasts (pIC_{50}). The molecular descriptors used to correlate the chemical structures with the respective pIC_{50} are based on multivariate image analysis (MIA) (Barigye and Freitas, 2016). These descriptors have proven to be efficient parameters in describing the chemical structures of a variety of datasets for modelling purposes; successful models involving physical, pharmacological and agrochemical properties have been obtained so far (see Daré et al., 2017; Goodarzi and Freitas, 2008; Pereira and Freitas, 2021, to name a few). Although 3D-QSAR is widely used for these purposes and even announced as QSAR web servers (Ragno, 2019; Wang et al., 2021), the conformational search and alignment processes required in 3D-QSAR are not straightforward. Moreover, the actual relevance of conformational information for QSAR purposes has been revisited (Daré et al., 2019; Daré et al., 2020; Daré and Freitas, 2021).

The so-called MIA-QSAR method has evolved from simple wireframe representation of molecules (Freitas et al., 2005) to colored ball-and-stick models (Barigye and Freitas, 2016; Nunes and Freitas, 2013). The latter provides a more realistic view of a molecule and can indeed incorporate chemical information, as follows. The MIA descriptors are digital units corresponding to pixels forming the image of a molecule and may be numbered according to the RGB system of colors, whose combination of the three channels (red, green, and blue) may give values that range from zero (black) to 765 (white). Considering that atoms in a molecule can be represented as circles with sizes proportional to the corresponding Van der Waals radii and that MIA descriptors vary both in value and orientation as the chemical structure changes, then topological and atomic properties can be encoded. Therefore, the chemical features ruling the herbicidal activities can be rationalized and the design of new agrochemical candidates may be guided by these outcomes. In addition, the grouping of two families of congeneric compounds increases the molecular diversity and agrochemical candidacy of derivatives, which can be particularly useful to control resistant weeds.

2. MATERIALS AND METHODS

Two series of benzamide and thiobenzamide derivatives were obtained from three different sources (Honda et al.,1990; Imramovsky et al., 2011; Sharma, 2016) and grouped

into a single data set. This grouping was possible because a) the herbicidal activities (pIC_{50}) were obtained using similar measurement methods, *i.e.* using chloroplasts isolated from spinach leaves (*Spinacia oleracea*), and b) the molecules share a common scaffold, *i.e.* this set is a congeneric series of compounds. 80 out of 86 compounds have shown accurate pIC_{50} values and were then used in the MIA-QSAR modeling (Table S1, Supplementary Material); data for 6 compounds are given as < 5 and, therefore, were not used. The MIA descriptors were generated from chemical images represented as wireframe sketches (conventional MIA-QSAR) and balls-and-sticks (augmented MIA-QSAR). The wireframe-based images were built using the ChemSketch program (ACD/ChemSketch, 2018), whereas the balls (atoms with sizes proportional to the Van der Waals radii) and sticks (chemical bonds) were designed using the GaussView program (Dennington et al., 2008). The images had to be superimposable for alignment purposes, since the congeneric moiety should not vary and only substituent changes are expected to affect the biological response (Figure S1, Supplementary Material). Since the color pixels may be renumbered, the atoms represented as circles were valued to make a correspondence with the respective Van de Waals radii (r_{vdW}), electronegativity values (ϵ), and the r_{vdW}/ϵ ratio (Table S2, Supplementary Material). These parameters give insight into hydrophobic, steric, and dipolar interactions that rule a ligand-enzyme fit. Consequently, four MIA-QSAR models were built for prediction purposes, namely the $\text{MIA}_{\text{sketch}}$, $\text{MIA}_{r_{\text{vdW}}}$, MIA_{ϵ} , and $\text{MIA}_{r_{\text{vdW}}/\epsilon}$ models.

Since the general procedure for building a MIA-QSAR model has been described in detail elsewhere (Freitas, 2006; Nunes and Freitas, 2013; Freitas et al., 2015), only a brief description is given herein, as follows. Each bitmap image was saved as a pixel matrix with a predefined $x \times y$ dimension and then the 80 images were grouped to yield an $80 \times x \times y$ three-way array. The three-way array was unfolded to a descriptor data matrix of 80 rows (corresponding to the 80 molecules) and $x \times y$ columns (corresponding to the MIA descriptors). This \mathbf{X} -matrix was split into training and test subsets (3:1 ratio, according to Tropsha (2010)) using Kennard-Stone sampling and the former was regressed against the y -block of pIC_{50} values through partial least squares (PLS). This procedure is illustrated in Figure S2, Supplementary Material. The PLS regression coefficients were used to predict the pIC_{50} data for the test set samples as an external validation procedure. The MIA-QSAR models were also internally validated employing leave-one-out cross-validation. The chance correlation was evaluated using a y -randomization test with ten repetitions, which gives the statistical difference between the correlations obtained from the real calibration and that simulated with the intact \mathbf{X} -matrix but scrambled y -block. The quality of the calibration and

validation was assessed by the root mean square errors (RMSE) and determination coefficients of the experimental *vs.* calculated pIC₅₀ plots, which correspond to r^2 for calibration, q^2 for leave-one-out cross-validation, and r^2_{pred} for external validation (Todeschini et al., 2020). The statistical relevance of $r^2_{y\text{-rand}}$ and r^2_{pred} was assessed by the $^c r^2_p$ and r^2_m parameters, respectively (Mitra et al., 2010; Roy et al., 2013). All these procedures and analyses were performed using the Chemoface program (Nunes et al., 2012). Further analyses were carried out to characterize the interpolation space encompassing the samples and to assure the robustness, reliability, and prediction capability of the models. These were assessed through an applicability domain test (Roy et al., 2015) (by plotting sample leverages *versus* Student's residuals) and additional validation strategies: bootstrapping validation (Kiralj and Ferreira, 2009), mean absolute error (MAE) (Roy et al., 2016), slopes of regression lines (k and k') (Golbraikh and Tropsha, 2002), concordance correlation coefficient (CCC) (Chirico and Gramatica, 2011), and different metrics that account for the difference between experimental and predicted data (Q^2_{Fn}) (Chirico and Gramatica, 2011). The criteria for acceptance of the models are: $r^2 > 0.6$; q^2 , $^c r^2_p$, r^2_p , r^2_m , and $Q^2_{\text{Fn}} > 0.5$; $0.85 < k$ and $k' < 1.15$; $\text{CCC} > 0.85$. Finally, the chemical contributions to the herbicidal behavior were assessed utilizing the MIA-plots (applied to augmented MIA-QSAR), which correspond to molecular contour maps describing the substituents that most affect (through variable importance in projection scores – VIP) and how they affect (through the PLS regression coefficients – \mathbf{b}) the biological response (Barigye et al., 2016).

A molecular docking study was carried out to explain the binding affinity of the (thio)benzamide herbicides against the PSII protein. The structure of *Spinacia oleracea* PSII-LHCII supercomplex used in the study was retrieved from the Protein Data Bank (PDB code: 3JCU) and obtained by electron microscopy with a resolution of 3.20 Å (Wei et al., 2016). The *S. oleracea* is the same species used to obtain the biological data. The 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 prepared ligands were docked into the photosystem II D1 active site (psbA, EC:1.10.3.9) using the Molegro Virtual Docker software (Thomsen and Christensen, 2006), a program for predicting the most likely binding conformation of a ligand towards a macromolecule (Cunha et al. 2011; Ramalho et al., 2012).

3. RESULTS

To check whether compounds are within the chemical space encompassing the molecular descriptors and modeled response, William's plots based on sample leverages and Student's residuals were obtained and no outliers could be unequivocally detected. Therefore, the whole series of compounds were used to build the MIA-QSAR models based on wireframes and GaussView images (MIA_{sketch} , MIA_{rVdW} , MIA_{ϵ} , and $MIA_{\text{rVdW}/\epsilon}$). The number of latent variables (PLS components) used to build the MIA-QSAR models described earlier was chosen based on the decay in the RMSECV and all these models achieved the necessary conditions to be considered as reliable and predictive, i.e. $r^2 > 0.6$, and q^2 and $r^2_{\text{pred}} > 0.5$, in addition to other parameters (see Table 1 for the statistical parameters and Figure 2 for the experimental \times calculated pIC_{50} plots). These models were externally validated using a bootstrapping procedure with 20 cycles, and the average measures are given in Table S3, Supplementary Material, and discussed in the next section.

Table 1. Statistical evaluation of the MIA-QSAR models using Kennard-Stone sampling to split the data set into training and test subsets.

Parameter	MIA_{sketch}	MIA_{rVdW}	MIA_{ϵ}	$MIA_{\text{rVdW}/\epsilon}$
PLS components	5	4	4	4
r^2	0.910	0.915	0.909	0.921
RMSEC	0.339	0.362	0.369	0.364
$r^2_{\text{y-rand}}$	0.599	0.457	0.493	0.479
$^c r^2_{\text{p}}$	0.532	0.648	0.615	0.638
$\text{RMSE}_{\text{y-rand}}$	0.716	0.912	0.869	0.934
q^2	0.742	0.826	0.801	0.845
RMSECV	0.581	0.520	0.548	0.512
r^2_{pred}	0.911	0.805	0.832	0.702
r^2_{m}	0.834	0.662	0.739	0.415
RMSEP	0.490	0.596	0.549	0.606
MAE	0.077	0.098	0.132	0.084
K	0.987	0.982	0.974	0.987
k'	1.007	1.008	1.018	1.003
Q^2_{F1}	0.912	0.825	0.866	0.775

Q^2_{F2}	0.909	0.797	0.813	0.696
Q^2_{F3}	0.813	0.769	0.798	0.782
CCC	0.953	0.893	0.907	0.817

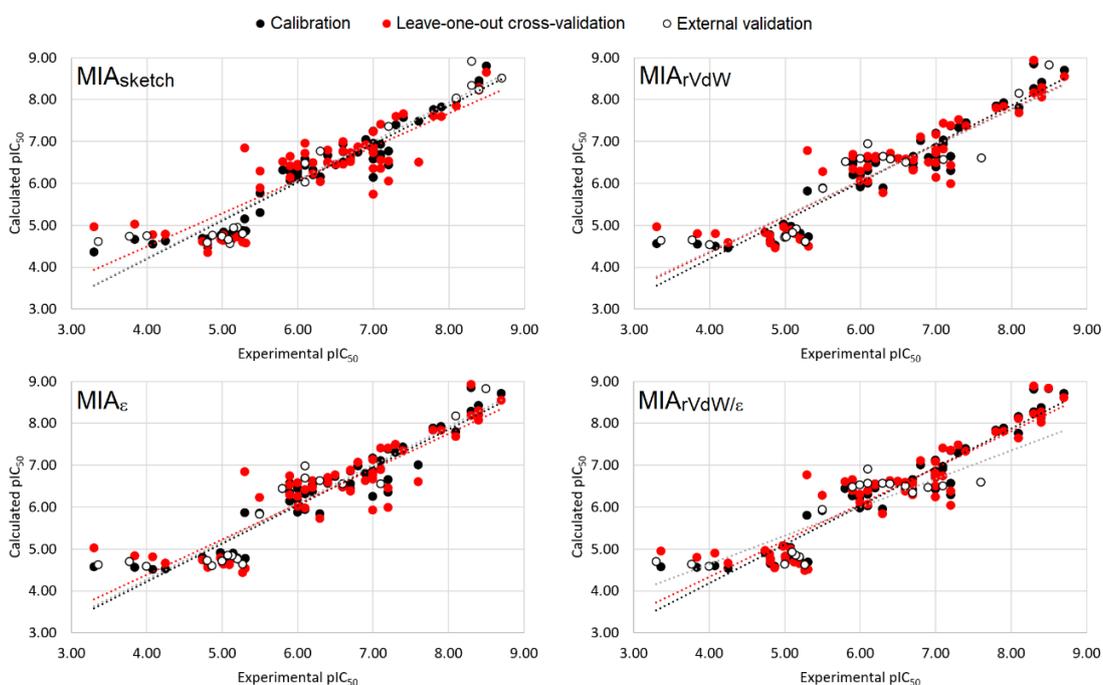


Figure 2. Correlation plots of experimental \times calculated pIC₅₀ for the four MIA-QSAR models.

MIA-plots were generated from the descriptors obtained in the GaussView chemical structures to interpret the chemical features affecting the pIC₅₀ values (Figure 3). The substituents that most influence the models are R¹, R², and R⁴. According to the **b**-plots, the chemical groups that most positively affect the pIC₅₀ values are X = S, R¹ = C2–C9 chain (mainly C5–C7 chain), R³ = OH, R⁴ = NO₂; a more subtle effect is observed for R² = CONHC5–C9. In turn, the substituents that most negatively affect the biological data are R¹ = 3-PhCl, R¹ = C14–C15 chain, R² = CONHEt, and R⁴ = Cl.

The outcomes from this analysis, as well as using a combination of chemical moieties of the most promising compounds of Table S1, guided the proposition of novel herbicide candidates, **P1–P18** (Table 2). **P1** and **P2** demonstrated to be the most promising agrochemical candidates, with predicted pIC₅₀ of 8.67 and 8.29, respectively. Their mechanism of action was evaluated through docking studies, particularly using the docking scores of the ligand-enzyme interactions (Supplementary Material), and the most relevant intermolecular interactions present in the binding pocket of the PSII enzyme (Figure 4). A

detailed analysis of the interactions, as well as a comparison with the QSAR results and with the experimentally known compounds, are given in the next section.

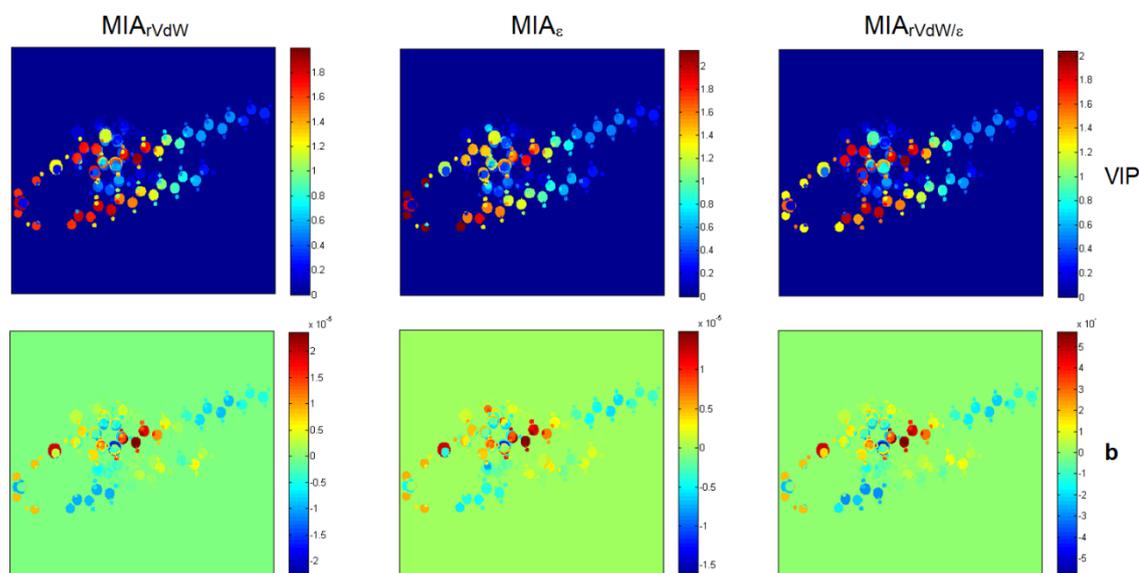


Figure 3. MIA-plots based on variable importance in projection (VIP) scores and PLS regression coefficients (**B**) for the three augmented MIA-QSAR models.

Table 2. Proposed (thio)benzamide herbicides and the respective calculated pIC₅₀ values (mean values for the four MIA-QSAR models).

Proposal	X	R ¹	R ²	R ³	R ⁴	R ⁵	pIC ₅₀
P1	S	Undecyl	H	OH	NO ₂	OH	8.67 ± 0.02
P2	S	Tridecyl	H	OH	NO ₂	OH	8.29 ± 0.04
P3	S	Nonyl	CONHPent	OH	NO ₂	OH	8.03 ± 0.44
P4	S	Nonyl	CONHHex	OH	NO ₂	OH	8.02 ± 0.39
P5	S	Decyl	CONHPent	OH	NO ₂	OH	8.01 ± 0.45
P6	S	Decyl	CONHHex	OH	NO ₂	OH	8.00 ± 0.40
P7	S	Decyl	CONHHept	OH	NO ₂	OH	7.98 ± 0.35
P8	S	Nonyl	CONHHept	OH	NO ₂	OH	7.94 ± 0.25
P9	S	Undecyl	CONHHex	OH	NO ₂	OH	7.91 ± 0.45
P10	S	Octyl	CONHPent	OH	NO ₂	OH	7.90 ± 0.43
P11	S	Undecyl	CONHPent	OH	NO ₂	OH	7.90 ± 0.47
P12	S	Octyl	CONHHex	OH	NO ₂	OH	7.89 ± 0.38
P13	S	Undecyl	CONHHept	OH	NO ₂	OH	7.88 ± 0.39
P14	S	Octyl	CONHHept	OH	NO ₂	OH	7.88 ± 0.34

P15	O	Nonyl	CONHHept	OH	NO ₂	OH	7.48 ± 0.37
P16	O	Decyl	CONHHept	OH	NO ₂	OH	7.46 ± 0.38
P17	O	Octyl	CONHHept	OH	NO ₂	OH	7.40 ± 0.45
P18	O	Undecyl	CONHHept	OH	NO ₂	OH	7.40 ± 0.48

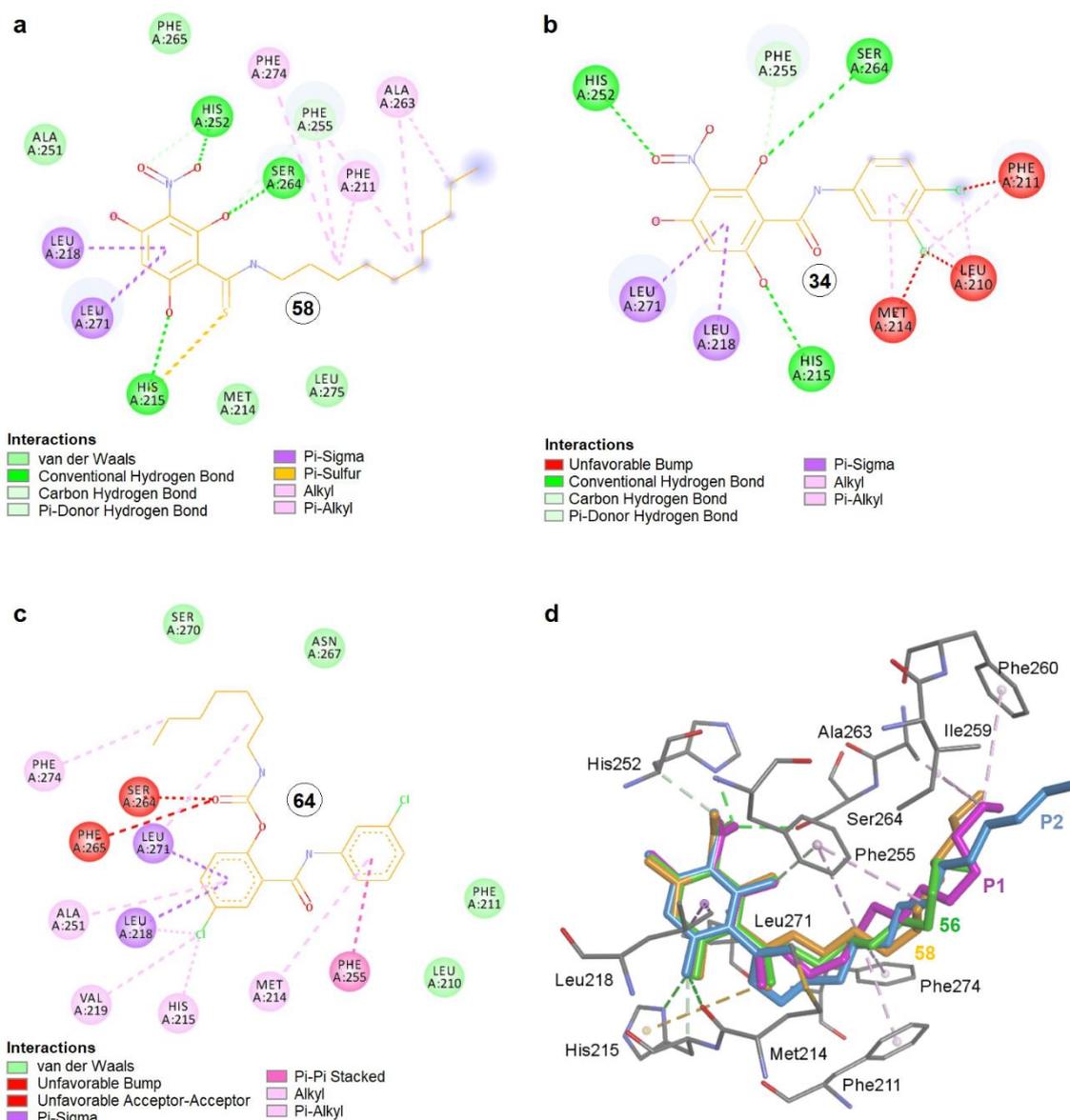


Figure 4. Important ligand-enzyme (PDB code: 3JCU) interactions obtained from the docking studies (the MolDock scores are given in the Supplementary Material).

4. DISCUSSION

Despite being obsolete, the conventional MIA-QSAR model based on wireframe sketches was used in further analysis, since it demonstrated a better prediction performance

($r^2_{\text{pred}} = 0.911$), while the augmented MIA-QSAR models are more chemically interpretable. This noticeable dichotomy between predictability and interpretability is indeed common and complimentary in QSAR studies (Fujita and Winkler, 2016). It is worth mentioning that no outlier compounds could be unequivocally detected from the applicability domain test, since slightly anomalous behavior was observed for only a few samples (in general, compounds **14**, **15**, **83**, and **84**), which vary with the descriptor type and whose removal does not affect the models significantly (Figure S3, Supplementary Material). These compounds possess either unique substituents ($R^1 =$ pentadecyl for **14** and octadecyl for **15**) or border pIC_{50} values (3.36 for **83** and 3.30 for **84**).

The only two statistical parameters below the cut-off values were r^2_{m} and CCC for the $\text{MIA}_{r_{\text{vdw}}/\epsilon}$ model, since the herbicidal activities for the test set were not regularly distributed along with the entire range of pIC_{50} values. Because the test set of compounds was chosen based on a Kennard-Stone sampling (Kennard and Stone, 1969), a second round of validation analyses was performed to evaluate the stability of the models upon a random selection of various test subsets (bootstrapping). Accordingly, highly predictive models were obtained (with all statistical parameters acceptable) and quite stable behavior was achieved for 20 cycles, *i.e.* the average values in general presented less than 20% of standard deviation in RMSEs and less than 10% in r^2 , q^2 , and r^2_{pred} (Table S3, Supplementary Material).

Considering that the MIA-QSAR models are reliable for the prediction of the pIC_{50} of new (thio)benzamide herbicides, the next step was to rationalize how the chemical structure affects the herbicidal activity in this series of compounds. This was accomplished by analyzing the variable importance in projection (VIP) scores, which describe the substituents that most explain the variance in the data set, and by the PLS regression coefficients (**b**), which inform on the MIA descriptors that influence either positively or negatively the herbicidal activity. Noticeably, according to the VIP contour maps of Figure 3, the MIA descriptors based on the Van der Waals radii and the Pauling's electronegativity, as well as the r_{vdw}/ϵ ratio, explain the data variance almost in the same manner, indicating that these properties are similarly important for herbicidal activity. Considering the guidance provided by the MIA-plots, *i.e.* the most promising substituents at the R^1 – R^4 positions, and also considering the combination of substructures of the most active (thio)benzamides present in each group of compounds (**1–59** and **60–86**), 18 herbicide candidates were proposed (**P1–P18**) and their pIC_{50} values were predicted using the four MIA-QSAR models (the mean values are given in Table 2 and those for each model is available in Table S4, Supplementary

Material). It is remarkable that a hydroxyl group rather than a carbamate at R² greatly favors the herbicidal potential, since **P2** and especially **P1** exhibited the highest calculated pIC₅₀ values, which are close to that of the best experimentally known derivative **56**. To attest whether these proposals are indeed promising, as well as to understand their mode of interaction with the biological target, a docking study was performed over the PSII complex available in the Protein Data Bank (PDB code: 3JCU).

The molecular docking results for the series of ligand compounds into the PSII D1 binding site highlight the key intermolecular interactions, which include hydrogen bonds, van der Waals, and electrostatic interactions. The key amino acid residues His215, Ser264, and Phe255 are involved in interactions with several ligand groups. The hydroxyl group at R² in compounds **1-59** is oriented towards the His215 residue to form a hydrogen bond; this is consistent with the QSAR finding that an OH group at R² positively influences the biological activity. Besides, the -NO₂ and -OH groups tend to interact with Ser264 and His252, respectively, through hydrogen bonds, whereas the benzamide ring interacts *via* π -sigma interactions with Leu218 and Leu 271 (see Figure 4a with **58** as benchmark ligand). It is clear from Table S1 (Supplementary Material) that the inhibitory activity values for the ligands **1-6** and **8-15** increase with the carbon-chain elongation at the R¹ substituent, in agreement with the QSAR finding that C2-C9 alkyl chains favor the bioactivity. These compounds are oriented to maximize van der Waals interactions with the hydrophobic amino acids Phe211, Phe255, Phe265, and Phe274, as well as with the plastoquinone cofactor. Replacement of linear alkyl chains with phenyl substituents (ligands **16-51**, **59**) causes a steric hindrance effect with closely positioned amino acids (Leu210, Phe211, and Met214; see Figure 4b with **34** as benchmark ligand); this effect explains the lower activity values for these ligands. A different conformation for the ligands **60-86** does not allow for hydrogen bond with His215, whereas the carboxyl group at R² can cause a steric hindrance effect with Phe265 and 264. The orientation of the 3-Cl-phenyl ring like in **64** induces π - π stacking interactions with Phe255, as shown in Figure 4c. Finally, the outcomes from the docking hint indicate that the proposed compounds **P1** (-129.2 kcal mol⁻¹) and **P2** (-135.4 kcal mol⁻¹) demonstrate binding affinities (MolDock scores) similar to that of ligands **56** (-117.9 kcal mol⁻¹, the most active compound) and **58** (-123.3 kcal mol⁻¹, the most stable similar compound in the pocket), as demonstrated in the QSAR prediction. This implies that our candidate scaffolds **P1** and **P2** have good stabilities in the active site of PSII D1 compared to the ligand molecules **56** and **58**. **P1**, **P2**, **56**, and **58** established hydrogen bonds with Ser264, His252, and His215; they can also establish π - π stacking interactions with Phe255 (Figure 4d).

5. CONCLUSIONS

Two sets of congeneric compounds were gathered, then yielding an enlarged data set with 80 (thio)benzamides, which provided consistent and reliable MIA-QSAR models. The MIA descriptors explained topochemical, electronegativity and atomic radius properties. Based on MIA-plots which map the molecular moieties that most affect the PSII inhibitory activity, as well as by mixing the substructures of the most promising compounds within the data set, a series of (thio)benzamide herbicide candidates were proposed. According to the prediction models, the best proposal is most likely a thiobenzamide derivative containing C11 aliphatic chain, nitro, and hydroxyl groups as substituents. The role of these substituent effects was rationalized through docking studies, which revealed important ligand-enzyme interactions. In addition, the intermolecular interaction energies obtained from the docking analysis are in reasonable agreement with the predicted pIC_{50} , thus providing further validation of the MIA-QSAR models. Finally, these findings may be useful to design new lead compounds with improved herbicidal performance, especially against resistant weeds.

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: 301371/2017-2), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

REFERENCES

- ACD/ChemSketch 2017.2.1 (2018). Advanced Chemistry Development, Inc., Toronto.
- 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://dx.doi.org/10.1039/c6ra09593c>.
- Barigye, S.J., Freitas, M.P., 2016. Ten years of the MIA-QSAR strategy: Hystorical development and applications. *Int. J. Quant. Struct.-Prop. Relat.*, 1, 64-77. <https://dx.doi.org/10.4018/ijqspr.2016010103>.
- Brookes, G., Barfoot, P., 2020. GM crop technology use 1996-2018: Farm income and production impacts. *GM Crops and Food*, 11, 242-261. <https://dx.doi.org/10.1080/21645698.2020.1779574>.
- Chirico, N., Gramatica, P., 2011. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.*, 51, 2320-2335. <https://dx.doi.org/10.1021/ci200211n>.
- da Cunha, E.F.F., Mancini, D.T., Ramalho, T.C., 2012. Molecular modeling of the *Toxoplasma gondii* adenosine kinase inhibitors. *Med. Chem. Res.*, 21, 590-600. <https://dx.doi.org/10.1007/s00044-011-9554-z>.
- Daré, J.K., Freitas, M.P., 2021. Different approaches to encode and model 3D information in a MIA-QSAR perspective. *Chemom. Intell. Lab. Sys.*, 212, 104286. <https://dx.doi.org/10.1016/j.chemolab.2021.104286>.
- Daré, J.K., Ramalho, T.C., Freitas, M.P., 2019. 3D perspective into MIA-QSAR: A case for anti-HCV agents. *Chem. Biol. Drug Des.*, 93, 1096-1104. <https://dx.doi.org/10.1111/cbdd.13440>.
- Daré, J.K., Silva, C.F., Freitas, M.P., 2017. Revealing chemophoric sites in organophosphorus insecticides through the MIA-QSPR modeling of soil sorption data. *Ecotox. Environ. Saf.*, 144, 560-563. <https://dx.doi.org/10.1016/j.ecoenv.2017.06.072>.
- Daré, J.K., Silva, D.R., Ramalho, T.C., Freitas, M.P., 2020. Conformational fingerprints in the modelling performance of MIA-QSAR: a case for SARS-CoV protease inhibitors. *Mol. Simul.*, 46, 1055-1061. <https://dx.doi.org/10.1080/08927022.2020.1800691>.

- Dennington, R.D., Keith, T.A., Millam, M.J., 2008. GaussView 5.0, Wallingford, CT.
- Devine, M., Duke, S.O., Fedtke, C., 1993. *Physiology of Herbicide Action*, Prentice Hall, New Jersey.
- Freitas M.P. (2006). MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzotrioles and their thio and sulfinyl congeners. *Org. Biomol. Chem.*, 4, 1154-1159. <https://dx.doi.org/10.1039/b516396j>.
- Freitas, M.P., Brown, S.D., Martins, J.A., 2005. MIA-QSAR: a simple 2D image-based approach for quantitative structure–activity relationship analysis. *J. Mol. Struct.*, 738, 149-154. <https://dx.doi.org/10.1016/j.molstruc.2004.11.065>.
- 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://dx.doi.org/10.1039/c4ra12070a>.
- Fujita, T., Winkler, D.A., 2016. Understanding the roles of the “two QSARs”. *J. Chem. Inf. Model.*, 56, 269–274. <https://dx.doi.org/10.1021/acs.jcim.5b00229>.
- Golbraikh, A., Tropsha, A., 2002. Beware of q^2 ! *J. Mol. Graph. Modell.*, 20, 269-276. [https://dx.doi.org/10.1016/s1093-3263\(01\)00123-1](https://dx.doi.org/10.1016/s1093-3263(01)00123-1).
- Goodarzi, M., Freitas, M.P., 2008. Predicting boiling points of aliphatic alcohols through multivariate image analysis applied to quantitative structure-property relationships. *J. Phys. Chem. A*, 112, 11263-11265. <http://dx.doi.org/10.1021/jp8059085>.
- Honda, I., Yoneyama, K., Iwamura, H., Konnai, M., Takahashi, N., Yoshida, S., 1990. Structure-activity relationship of 3-nitro-2,4,6-trihydroxybenzamide derivatives in photosynthetic electron transport inhibition. *Agric. Bio. Chem.*, 54, 1227-1233. <https://dx.doi.org/10.1080/00021369.1990.10870133>.
- Imramovsky, A., Pesko, M., Ferriz, J.M., Kralova, K., Vinsova, J., Jampilek, J., 2011. Photosynthesis—Inhibiting efficiency of 4-chloro-2-(chlorophenylcarbamoyl)phenyl alkylcarbamates. *Bioorg. Med. Chem. Lett.*, 21, 4564-4567. <https://dx.doi.org/10.1016/j.bmcl.2011.05.118>.
- Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. *Technometrics*, 11, 137-148.

- Kiralj, R., Ferreira, M.M.C., 2009. Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *J. Braz. Chem. Soc.*, 20, 770-787. <https://dx.doi.org/10.1590/s0103-50532009000400021>.
- 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://dx.doi.org/10.1080/08927022.2010.503326>.
- Nunes, C.A., Freitas, M.P., 2013. Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors. *Eur. J. Med. Chem.*, 62, 297-300. <https://dx.doi.org/10.1016/j.ejmech.2013.01.005>.
- 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>.
- Oettmeier, W., 2003. Herbicides, inhibitors of photosynthesis at photosystem II. In J.R. Plimmer, N.R. Ragsdale, D.W. Gammon (Eds.), *Encyclopedia of Agrochemicals*, John Wiley and Sons, Hoboken.
- Pereira, I.V., Freitas, M.P., 2021. Double focus in the modelling of anti-influenza properties of 2-iminobenzimidazolines: pharmacology and toxicology. *SAR QSAR Environ. Res.*, 32, 643-654. <https://dx.doi.org/10.1080/1062936x.2021.1950832>.
- Ragno, R. (2019). www.3d-qsar.com: a web portal that brings 3-D QSAR to all electronic devices-the Py-CoMFA web application as tool to build models from pre-aligned datasets. *J. Comput.-Aid. Mol. Des.*, 33, 855-864. <https://dx.doi.org/10.1007/s10822-019-00231-x>.
- Ramalho, T.C., França, T.C.C., Rennó, M.N., Guimarães, A.P., da Cunha, E.F.F., Kuca, K., 2012. Development of new acetylcholinesterase reactivators: Molecular modeling versus in vitro data. *Chem.-Biol. Int.*, 185, 73-77. <https://dx.doi.org/10.1016/j.cbi.2010.05.016>.
- Roy, K., Chakraborty, P., Mitra, I., Ojha, P.K., Kar, S., Das, R.N., 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://dx.doi.org/10.1002/jcc.23231>.
- Roy, K., Das, R.N., Ambure, P., Aher, R.B., 2016. Be aware of error measures. Further studies on validation and predictive QSAR models. *Chemom. Intell. Lab. Sys.*, 152, 18-33. <https://dx.doi.org/10.1016/j.chemolab.2016.01.008>.

- 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://dx.doi.org/10.1016/j.chemolab.2015.04.013>.
- Sharma M.C., 2016. Identification of 3-nitro-2,4,6-trihydroxybenzamide derivatives as photosynthetic electron transport inhibitors by QSAR and pharmacophore studies. *Interdiscip. Sci. Comput. Life Sci.*, 8, 109-121. <https://dx.doi.org/10.1007/s12539-015-0019-9>.
- Storkey, J., Bruce, T.J.A., McMillan V.E., Neve, P., 2019. The future of sustainable crop protection relies on increased diversity of cropping systems and landscapes. In G. Lemaire, S. Kronberg, P.C.F Carvalho, S. Recous (Eds.), *Agroecosystem Diversity* (pp. 199-209). Elsevier B.V., Amsterdam,
- Todeschini, R., Consonni, V., Ballabio, D., Grisoni, F., 2020. Chemometrics for QSAR Modeling. In S.D. Brown, R. Tauler, B. Walczak (Eds.), *Comprehensive Chemometrics* (pp. 599-634). Elsevier, Amsterdam.
- Thomsen, R., Christensen, M.H., 2006. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.*, 49, 3315-3321. <https://dx.doi.org/10.1021/jm051197e>.
- Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.*, 29, 476-488. <https://dx.doi.org/10.1002/minf.201000061>.
- Wang, Y-L., Wang, F., Shi, X-X., Jia, C-Y., Wu, F-X., Hao, G-F., Yang, G-F., 2021. Cloud 3D-QSAR: a web tool for the development of quantitative structure–activity relationship models in drug discovery. *Brief. Bioinf.*, 22, bbaa276. <https://doi.org/10.1093/bib/bbaa276>.
- Wei, X., Su, X., Cao, P., Liu, X., Chang, W., Li, M., Zhang, X., Liu, Z., 2016. Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. *Nature*, 534, 69-74. <https://dx.doi.org/10.1038/nature18020>.

APPENDIX A - SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIAL

MIA-QSAR study of the structural merging of (thio)benzamide herbicides with photosynthetic system II inhibitory activities

Ingrid V. Pereira,^a Joyce K. Daré,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

^a Departamento de Química, Instituto de Ciências Naturais, Universidade Federal de Lavras, 37200-900, Lavras, MG, Brazil. E-mail: matheus@ufla.br

Page 71. **Table S1.** Benzamide (X = O) and thiobenzamide (X = S) derivatives, and the respective pIC₅₀ values used in the MIA-QSAR modeling.^a

Page 73. **Table S2.** Original RGB pixel values and descriptor values proportional to atomic properties assigned to the atoms present in the molecules of the data set.

Page 74. **Figure S1.** Superposed images showing how the (thio)benzamide derivatives are represented and vary in a congeneric series according to the conventional and augmented MIA-QSAR approaches.

Page 74. **Figure S2.** General workflow used to build the MIA-QSAR models.

Page 75. **Figure S3.** Applicability domain tests for the four MIA-QSAR data sets.

Page 76. **Table S3.** Statistical results of robustness and stability obtained from the bootstrapping experiments (mean values for 20 cycles).

Page 77. **Table S4.** Predicted pIC₅₀ values for the proposed compounds **P1–P18** using the four MIA-QSAR models.

Page 78. **Figure S4.** MolDock scores (kcal mol⁻¹) for the ligands of the known and proposed benzamide herbicides. Highlights: red = seeming outliers (**14** and **15**); green = best experimental benzamide herbicide (**56**); yellow = most stable analog of **56** in the pocket (**58**); magenta = **P1**; gray = **P2**.

Table S1. Benzamide (X = O) and thiobenzamide (X = S) derivatives, and the respective pIC₅₀ values used in the MIA-QSAR modeling.^a

Cpd	X	R ¹	R ²	R ³	R ⁴	R ⁵	pIC ₅₀
1	O	Methyl	H	OH	NO ₂	OH	5.50
2	O	Ethyl	H	OH	NO ₂	OH	6.10
3	O	Propyl	H	OH	NO ₂	OH	6.00
4	O	Butyl	H	OH	NO ₂	OH	6.10
5	O	Pentyl	H	OH	NO ₂	OH	7.00
6	O	Hexyl	H	OH	NO ₂	OH	7.20
7	O	Cyclohexyl	H	OH	NO ₂	OH	6.20
8	O	Heptyl	H	OH	NO ₂	OH	7.80
9	O	Octyl	H	OH	NO ₂	OH	8.10
10	O	Nonyl	H	OH	NO ₂	OH	8.30
11	O	Decyl	H	OH	NO ₂	OH	8.40
12	O	Undecyl	H	OH	NO ₂	OH	8.40
13	O	Tridecyl	H	OH	NO ₂	OH	8.10
14	O	Pentadecyl	H	OH	NO ₂	OH	7.30
15	O	Octadecyl	H	OH	NO ₂	OH	7.10
16	O	Benzyl	H	OH	NO ₂	OH	5.50
17	O	2-Phenetyl	H	OH	NO ₂	OH	5.30
18	O	(<i>R</i>)-1-Phenetyl	H	OH	NO ₂	OH	6.00
19	O	(<i>S</i>)-1-Phenetyl	H	OH	NO ₂	OH	6.30
20	O	3-Phenylpropyl	H	OH	NO ₂	OH	6.80
21	O	4-Phenylbutyl	H	OH	NO ₂	OH	7.40
22	O	2-Phenoxyethyl	H	OH	NO ₂	OH	6.10
23	O	2-(4-Cl-Phenoxy)ethyl	H	OH	NO ₂	OH	7.10
24	O	Phenyl	H	OH	NO ₂	OH	6.30
25	O	<i>N</i> -Me-Phenyl	H	OH	NO ₂	OH	^b
26	O	2-Cl-Phenyl	H	OH	NO ₂	OH	5.90
27	O	3-Cl-Phenyl	H	OH	NO ₂	OH	6.70
28	O	4-Cl-Phenyl	H	OH	NO ₂	OH	7.10
29	O	4-Br-Phenyl	H	OH	NO ₂	OH	7.20
30	O	2-F-Phenyl	H	OH	NO ₂	OH	6.00
31	O	3-F-Phenyl	H	OH	NO ₂	OH	6.60
32	O	4-F-Phenyl	H	OH	NO ₂	OH	6.40
33	O	3,5-Cl ₂ -Phenyl	H	OH	NO ₂	OH	6.70
34	O	3,4-Cl ₂ -Phenyl	H	OH	NO ₂	OH	7.20
35	O	2,4-Cl ₂ -Phenyl	H	OH	NO ₂	OH	7.00
36	O	2,3-Cl ₂ -Phenyl	H	OH	NO ₂	OH	5.90
37	O	2-CF ₃ -Phenyl	H	OH	NO ₂	OH	^b
38	O	3-CF ₃ -Phenyl	H	OH	NO ₂	OH	6.90
39	O	4-CF ₃ -Phenyl	H	OH	NO ₂	OH	7.60
40	O	2-Me-Phenyl	H	OH	NO ₂	OH	5.80
41	O	3-Me-Phenyl	H	OH	NO ₂	OH	6.60
42	O	4-Me-Phenyl	H	OH	NO ₂	OH	6.10
43	O	4-Et-Phenyl	H	OH	NO ₂	OH	6.50

44	O	4- <i>i</i> Pr-Phenyl	H	OH	NO ₂	OH	6.70
45	O	4-Pr-Phenyl	H	OH	NO ₂	OH	7.00
46	O	2-MeO-Phenyl	H	OH	NO ₂	OH	6.20
47	O	3-MeO-Phenyl	H	OH	NO ₂	OH	6.40
48	O	4-MeO-Phenyl	H	OH	NO ₂	OH	5.90
49	O	3,4-(MeO) ₂ -Phenyl	H	OH	NO ₂	OH	^b
50	O	4-NMe ₂ -Phenyl	H	OH	NO ₂	OH	^b
51	O	4-NO ₂ -Phenyl	H	OH	NO ₂	OH	6.70
52	S	Ethyl	H	OH	NO ₂	OH	6.60
53	S	Butyl	H	OH	NO ₂	OH	7.00
54	S	Hexyl	H	OH	NO ₂	OH	7.90
55	S	Heptyl	H	OH	NO ₂	OH	8.40
56	S	Octyl	H	OH	NO ₂	OH	8.70
57	S	Nonyl	H	OH	NO ₂	OH	8.50
58	S	Decyl	H	OH	NO ₂	OH	8.30
59	S	Phenyl	H	OH	NO ₂	OH	7.00
60	O	3-Cl-Phenyl	CONHEt	H	Cl	H	4.81
61	O	3-Cl-Phenyl	CONHBu	H	Cl	H	4.87
62	O	3-Cl-Phenyl	CONHPent	H	Cl	H	5.27
63	O	3-Cl-Phenyl	CONHHex	H	Cl	H	5.27
64	O	3-Cl-Phenyl	CONHHept	H	Cl	H	5.31
65	O	3-Cl-Phenyl	CONHOct	H	Cl	H	5.00
66	O	3-Cl-Phenyl	CONHNon	H	Cl	H	4.81
67	O	3-Cl-Phenyl	CONHDec	H	Cl	H	4.80
68	O	3-Cl-Phenyl	CONHUndec	H	Cl	H	4.74
69	O	4-Cl-Phenyl	CONHEt	H	Cl	H	5.01
70	O	4-Cl-Phenyl	CONHBu	H	Cl	H	5.02
71	O	4-Cl-Phenyl	CONHPent	H	Cl	H	5.20
72	O	4-Cl-Phenyl	CONHHex	H	Cl	H	5.20
73	O	4-Cl-Phenyl	CONHHept	H	Cl	H	5.15
74	O	4-Cl-Phenyl	CONHOct	H	Cl	H	5.13
75	O	4-Cl-Phenyl	CONHNon	H	Cl	H	5.10
76	O	4-Cl-Phenyl	CONHDec	H	Cl	H	5.08
77	O	4-Cl-Phenyl	CONHUndec	H	Cl	H	4.98
78	O	3,4-Cl-Phenyl	CONHEt	H	Cl	H	4.08
79	O	3,4-Cl-Phenyl	CONHBu	H	Cl	H	4.25
80	O	3,4-Cl-Phenyl	CONHPent	H	Cl	H	4.00
81	O	3,4-Cl-Phenyl	CONHHex	H	Cl	H	3.84
82	O	3,4-Cl-Phenyl	CONHHept	H	Cl	H	3.77
83	O	3,4-Cl-Phenyl	CONHOct	H	Cl	H	3.36
84	O	3,4-Cl-Phenyl	CONHNon	H	Cl	H	3.30
85	O	3,4-Cl-Phenyl	CONHDec	H	Cl	H	^b
86	O	3,4-Cl-Phenyl	CONHUndec	H	Cl	H	^b

^a Compounds **1-59** from Honda et al. (1990) and Sharma (2016), and compounds **60-86** from Imramovsky et al. (2011). ^b Inaccurate values not considered in the modeling.

Table S2. Original RGB pixel values and descriptor values proportional to atomic properties assigned to the atoms present in the molecules of the data set.

Atom/Chemical bond	Red	Green	Blue	RGB	rvaw	ϵ	rvaw/ϵ
Carbon	142	142	142	426	170	255	667
Hydrogen	204	204	204	612	120	220	545
Nitrogen	25	25	229	279	155	304	510
Oxygen	229	0	0	229	152	344	442
Sulfur	255	198	40	493	180	258	698
Fluorine	178	255	255	688	147	400	367
Chlorine	25	239	25	289	175	316	554
Bromine	165	33	33	231	185	296	625
Chemical bond	205	204	204	613	0.1	0.1	0.1

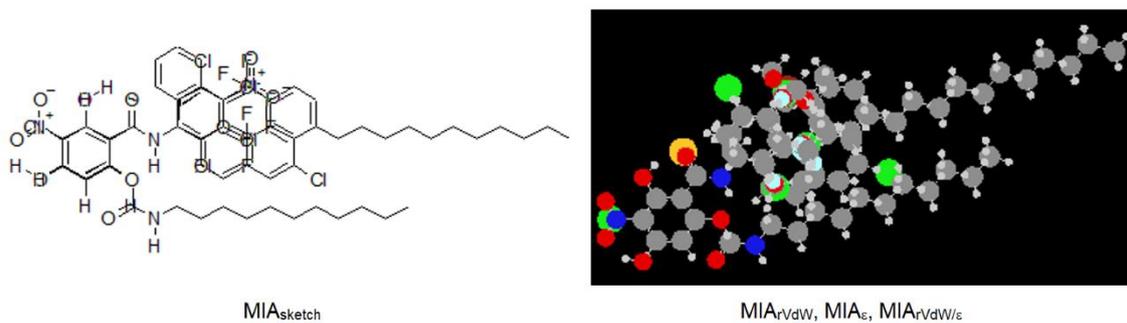


Figure S1. Superposed images showing how the (thio)benzamide derivatives are represented and vary in a congeneric series according to the conventional and augmented MIA-QSAR approaches.

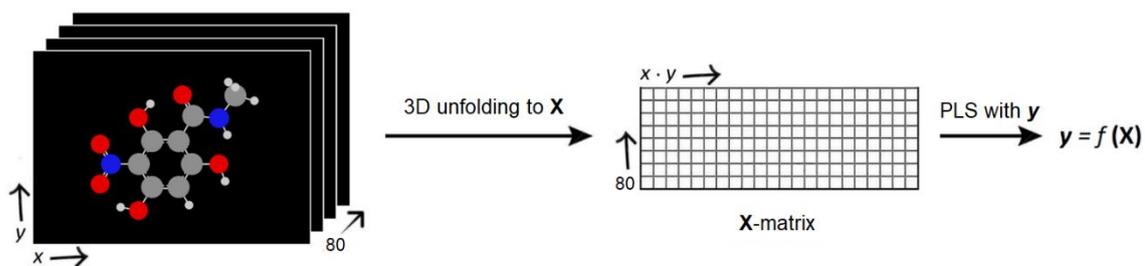


Figure S2. General workflow used to build the MIA-QSAR models.

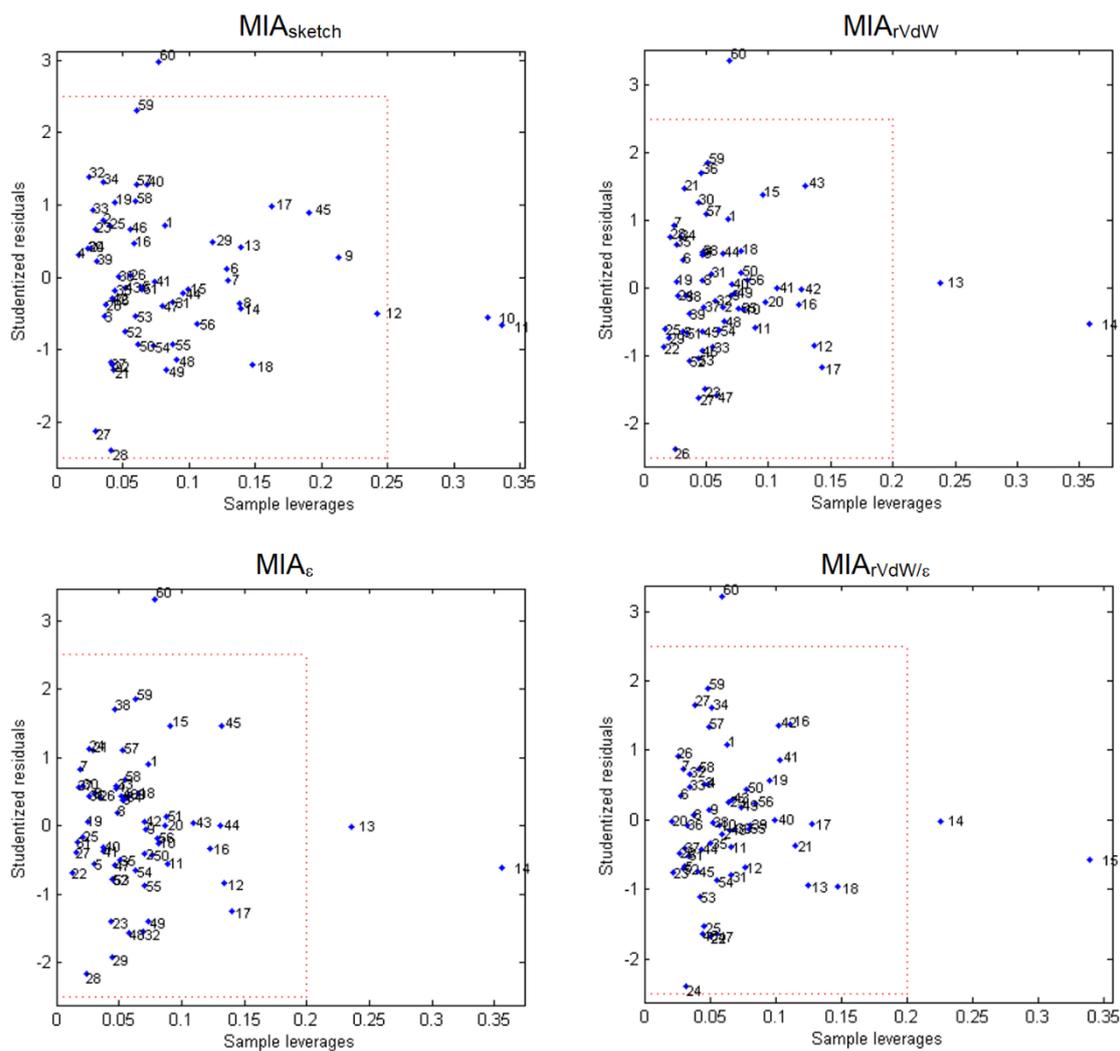


Figure S3. Applicability domain tests for the four MIA-QSAR data sets. The point labels do not correspond to the compound numbering, since the test set samples vary with the descriptor type; samples which are out of the cut-off values correspond to compounds **14**, **15**, and **84** (the point label 59 at the border line in the MIA_{sketch} plot corresponds to **83**).

Table S3. Statistical results of robustness and stability obtained from the bootstrapping experiments (mean values for 20 cycles).

Parameter	MIA _{sketch}	MIA _{rVdW}	MIA _{ϵ}	MIA _{rVdW/ϵ}
PLS components	4.7 \pm 0.7	3.9 \pm 0.7	3.8 \pm 0.7	3.9 \pm 0.7
r^2	0.916 \pm 0.016	0.898 \pm 0.015	0.884 \pm 0.049	0.895 \pm 0.012
RMSEC	0.370 \pm 0.035	0.409 \pm 0.033	0.416 \pm 0.045	0.415 \pm 0.027
$r^2_{y\text{-rand}}$	0.528 \pm 0.075	0.481 \pm 0.085	0.423 \pm 0.085	0.411 \pm 0.075
$c_r^2_p$	0.589 \pm 0.050	0.645 \pm 0.052	0.646 \pm 0.045	0.656 \pm 0.049
RMSE _{y-rand}	0.879 \pm 0.086	1.006 \pm 0.217	1.012 \pm 0.206	0.983 \pm 0.085
q^2	0.807 \pm 0.026	0.803 \pm 0.027	0.798 \pm 0.029	0.806 \pm 0.028
RMSECV	0.566 \pm 0.034	0.571 \pm 0.036	0.578 \pm 0.038	0.566 \pm 0.034
r^2_{pred}	0.843 \pm 0.070	0.839 \pm 0.066	0.835 \pm 0.064	0.821 \pm 0.105
r^2_m	0.740 \pm 0.115	0.744 \pm 0.126	0.741 \pm 0.124	0.745 \pm 0.131
RMSE _p	0.517 \pm 0.107	0.523 \pm 0.097	0.533 \pm 0.095	0.525 \pm 0.091
MAE	0.096 \pm 0.083	0.094 \pm 0.083	0.105 \pm 0.082	0.093 \pm 0.086
k	0.993 \pm 0.019	0.996 \pm 0.019	0.997 \pm 0.021	0.995 \pm 0.019
k'	1.001 \pm 0.018	0.998 \pm 0.018	0.996 \pm 0.020	0.998 \pm 0.018
Q^2_{F1}	0.820 \pm 0.085	0.817 \pm 0.080	0.812 \pm 0.079	0.818 \pm 0.077
Q^2_{F2}	0.810 \pm 0.092	0.807 \pm 0.086	0.802 \pm 0.085	0.808 \pm 0.083
Q^2_{F3}	0.831 \pm 0.062	0.827 \pm 0.059	0.822 \pm 0.056	0.827 \pm 0.055
CCC	0.907 \pm 0.045	0.905 \pm 0.042	0.902 \pm 0.041	0.905 \pm 0.043

Table S4. Predicted pIC₅₀ values for the proposed compounds **P1–P18** using the four MIA-QSAR models.

Proposal	MIA _{sketch}	MIA _{rVdW}	MIA _ε	MIA _{rVdW/ε}
P1	8.69	8.66	8.64	8.68
P2	8.34	8.27	8.26	8.29
P3	8.66	7.81	8.01	7.64
P4	8.57	7.83	8.01	7.67
P5	8.66	7.79	7.99	7.62
P6	8.57	7.81	7.99	7.64
P7	8.47	7.81	7.98	7.66
P8	8.26	7.83	8.00	7.68
P9	8.56	7.69	7.87	7.52
P10	8.51	7.69	7.88	7.52
P11	8.57	7.67	7.87	7.50
P12	8.42	7.71	7.89	7.55
P13	8.44	7.69	7.86	7.54
P14	8.35	7.71	7.88	7.56
P15	7.99	7.29	7.50	7.13
P16	7.99	7.27	7.48	7.11
P17	8.04	7.17	7.37	7.01
P18	8.09	7.15	7.36	6.99

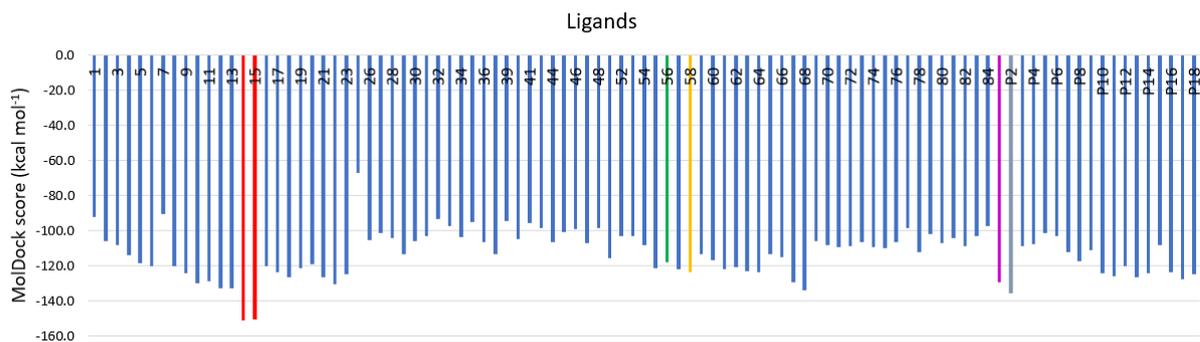


Figure S4. MolDock scores (kcal mol⁻¹) for the ligands of the known and proposed benzamide herbicides. Highlights: red = seeming outliers (**14** and **15**); green = best experimental benzamide herbicide (**56**); yellow = most stable analog of **56** in the pocket (**58**); magenta = **P1**; gray = **P2**.