



LUIZ DOS REIS CAPUCHO

**MODELAGEM MIA-QSAR E DOCKING MOLECULAR NA
PROPOSIÇÃO DE NOVAS TRICETONAS COMO
POTENCIAIS INIBIDORAS DA ENZIMA HPPD**

LAVRAS - MG

2023

LUIZ DOS REIS CAPUCHO

**MODELAGEM MIA-QSAR E DOCKING MOLECULAR NA PROPOSIÇÃO DE
NOVAS TRICETONAS COMO POTENCIAIS INIBIDORAS DA ENZIMA HPPD**

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

Prof. Dr. Matheus Puggina de Freitas

Orientador

LAVRAS – MG

2023

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

Capucho, Luiz dos Reis.

Modelagem MIA-QSAR e Docking Molecular na proposição
de novas tricetonas como potenciais inibidoras da enzima HPPD /
Luiz dos Reis Capucho. - 2023.

96 p. : il.

Orientador(a): Matheus Puggina de Freitas.

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

Bibliografia.

1. MIA-QSAR. 2. Docking Molecular. 3. Herbicida. I. Freitas,
Matheus Puggina de. II. Título.

LUIZ DOS REIS CAPUCHO

**MODELAGEM MIA-QSAR E DOCKING MOLECULAR NA PROPOSIÇÃO DE
NOVAS TRICETONAS COMO POTENCIAIS INIBIDORAS DA ENZIMA HPPD**

**MIA-QSAR MODELING AND MOLECULAR DOCKING TO PROPOSE NEW
TRIKETONES AS POTENTIAL HPPD ENZYME INHIBITORS**

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

APROVADA em 19 de janeiro de 2023.

Dr. João Eustáquio Antunes UFJF

Dr.^a Daiana Teixeira Mancini UFLA

Prof. Dr. Matheus Puggina de Freitas

Orientador

LAVRAS – MG

2023

*Aos meus pais Luzinete e Luiz Antônio, que muito contribuíram para a construção do que
hoje sou.*

Dedico

AGRADECIMENTOS

Agradeço à Deus toda força, liberdade e vitalidade; à Natureza pelos desafios e motivações; e às vidas presentes e passadas pela evolução que resultou na oferta dos recursos e ferramentas.

Agradeço ao meu pai e minha mãe, pelo apoio, atenção e exemplo de resiliência; às minhas irmãs, pelo incentivo, participação e auxílio; aos meus sobrinhos pela alegria e exemplo de esperança; e aos meus amigos pela torcida e momentos de descontração.

Agradeço ao meu orientador, por toda confiança, parceria e ensinamento; ao grupo de pesquisa, pelo acompanhamento e colaboração; à Universidade Federal de Lavras, pela oferta de estrutura e espaço; ao programa de pós-graduação em Agroquímica, pelo suporte e condições para o desenvolvimento deste trabalho; às agências de fomento CAPES, CNPq e FAPEMIG, pelo aporte e amparo na realização da pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

RESUMO

O problema de resistência de ervas tem causado significativa preocupação e orientado cada vez mais estudos referentes a compostos herbicidas. No objetivo de propor novos candidatos para superar esse problema recorrente na classe das tricetonas, dois conjuntos de moléculas foram submetidos a uma modelagem MIA-QSAR. Os descritores inicialmente obtidos em valores de pixels da escala RGB foram traduzidos a fim de considerar os efeitos de eletronegatividade (ϵ), raio de van der Waals (r_{vdW}) e a razão r_{vdW}/ϵ , e posteriormente submetidos ao método PLS de regressão, com validação realizada a partir de procedimento de randomização amostral. Os resultados obtidos em ambos os conjuntos demonstram que os três parâmetros avaliados contribuíram significativamente para a atividade observada dos compostos, com o conjunto de atividades em pIC_{50} apresentando melhores valores para o parâmetro r_{vdW}/ϵ , cuja média para 15 ciclos foi $R^2 = 0,823 \pm 0,022$, $Q^2 = 0,685 \pm 0,027$, $R^2_{pred} = 0,702 \pm 0,099$, sendo todos os valores aceitáveis por estarem acima de 0,5. No conjunto de atividades referentes a pK_i , apenas um modelo demonstrou ser preditivo, gerando valores estatísticos de $R^2 = 0,7180$, $Q^2 = 0,5090$, $R^2_{pred} = 0,7101$. Os MIA plot's **b** e **VIP** foram empregados para avaliar a influência dos substituintes sobre as variáveis de resposta e, subsequentemente, para a proposição de novos candidatos. Foram obtidas 8 e 14 estruturas com melhor atividade que o composto comercial padrão mesotriona para os conjuntos contendo atividades *in vitro* em pIC_{50} e pK_i respectivamente. Análises de *docking* molecular foram realizadas de modo complementar, visando a investigação de interações das moléculas propostas com o sítio ativo da enzima hidroxifenilpiruvato dioxigenase (HPPD), indicando que, para ambos os conjuntos, as propostas de melhor atividade apresentaram maior estabilidade no complexo enzimático do que a estrutura de maior atividade do conjunto total e que o composto padrão mesotriona. A síntese química dos compostos torna-se, portanto, promissora na validação das evidências teóricas apresentadas no presente trabalho.

Palavras-chave: Tricetonas. HPPD. Herbicida. MIA-QSAR. *Docking* molecular.

ABSTRACT

The problem of weed resistance has caused significant concern and has guided studies regarding herbicidal compounds. To propose new candidates to overcome this recurring problem in the class of triketones, two sets of molecules were subjected to MIA-QSAR modeling. The descriptors initially obtained in pixel values of the RGB scale were translated to consider the effects of electronegativity (ϵ), van der Waals radius (r_{vdW}) and the r_{vdW}/ϵ ratio, and subsequently submitted to the PLS regression method, with validation carried out from a sample randomization procedure. The results obtained in both sets demonstrate that the three evaluated parameters contributed significantly to the observed activity of the compounds, with the set of activities at pIC_{50} showing better values for the parameter r_{vdW}/ϵ , whose average for 15 cycles was $R^2 = 0.823 \pm 0.022$, $Q^2 = 0.685 \pm 0.027$, $R^2_{pred} = 0.702 \pm 0.099$, all acceptable values being above 0.5. In the set of activities related to pK_i , only one model proved to be predictive, generating statistical values of $R^2 = 0.7180$, $Q^2 = 0.5090$, $R^2_{pred} = 0.7101$. The MIA plot's **b** and **VIP** were used to evaluate the influence of substituents on the response variables and, subsequently, to propose new candidates; 8 and 14 structures with better activity than the standard commercial compound mesotrione were obtained for the sets containing *in vitro* activities in pIC_{50} and pK_i respectively. Molecular docking analyzes were carried out in a complementary way, aiming at investigating the interactions of the proposed molecules with the active site of the enzyme hydroxyphenylpyruvate dioxygenase (HPPD), indicating that, for both sets, the proposals with better activity presented greater stability in the enzymatic complex of the than the highest activity structure of the total pool and the standard compound mesotrione. The chemical synthesis of compounds becomes, therefore, promising in validating the theoretical evidence presented in this work.

Keywords: Triketones. HPPD. Herbicide. MIA-QSAR. Molecular docking.

SUMÁRIO

PRIMEIRA PARTE	10
1 INTRODUÇÃO.....	11
2 OBJETIVOS.....	14
2.1 <i>Objetivo geral</i>	14
2.2 <i>Objetivos específicos</i>	14
3 REFERENCIAL TEÓRICO	15
3.1 <i>Ervas daninhas e herbicidas</i>	15
3.2 <i>HPPD</i>	19
3.3 <i>Tricetonas</i>	30
3.4 <i>QSAR</i>	32
3.5 <i>Docking Molecular</i>	40
4 METODOLOGIA.....	45
4.1 <i>Softwares Utilizados</i>	45
4.2 <i>Métodos</i>	45
REFERÊNCIAS BIBLIOGRÁFICAS.....	48
SEGUNDA PARTE	56
Artigo 1 – Normas do periódico Journal of Science of Food and Agriculture (versão preliminar)	57
Artigo 2 – Normas do periódico Journal of Pest Science (versão preliminar)	77

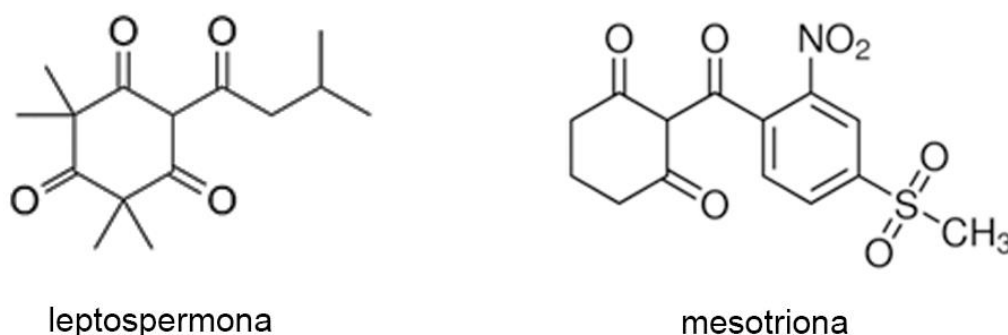
PRIMEIRA PARTE

1 INTRODUÇÃO

A facilidade de manejo e seletividade que os herbicidas apresentam os tornam grandes aliados de produtores agrícolas, permitindo colheitas de sucesso a custos e demandas de mão de obra relativamente baixos. Desse modo, pesquisas relacionadas ao desenvolvimento de novos princípios ativos ganham cada vez mais espaço, dados os benefícios do controle químico em detrimento a outras práticas de mitigação de pragas (ZIMDAHL, 2018a).

Dentre esses estudos, a mesotriona, princípio ativo do herbicida *Callisto*, foi patenteada pela empresa Syngenta e desenvolvida a partir do composto leptospermona. Isolada por Penfold em 1926, a leptospermona teve seu efeito alelopático conhecido apenas em meados do século 70 (FU *et al.*, 2019a), servindo posteriormente como molde estrutural ao composto comercial sintético cuja semelhança é verificada através da Figura 1.

Figura 1 - Representação estrutural da molécula leptospermona e mesotriona.



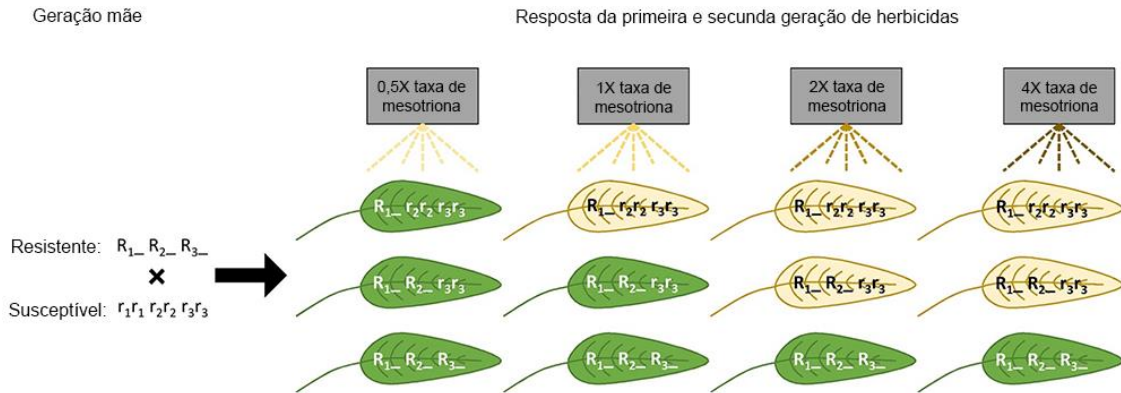
Fonte: Do autor (2022).

A aplicação da mesotriona em lavouras foi de grande aceitabilidade, sendo amplamente empregada em culturas de arroz, milho, soja e alguns outros cultivares (SUN *et al.*, 2012). O uso exaustivo desse composto, porém, resultou no problema da resistência decorrente da evolução genética das espécies, fazendo com que produtores agrícolas buscassem novos métodos de controle e adotassem práticas que, às vezes, apresentam prejuízos à saúde humana e ao meio ambiente.

Atualmente, já se conhece a resistência à herbicidas atuantes em 23 mecanismos de ação em um total de 26 conhecidos (KARAM *et al.*, 2018). Nesse contexto, o composto comercial tricetônico mais utilizado na inibição da enzima hidroxifenilpiruvato dioxigenase (HPPD), cuja ineficiência no tratamento de ervas já havia sido reportada (SILVA; MONQUERO; MUNHOZ, 2015), também sofreu com os efeitos evolutivos, trazendo ao meio acadêmico trabalhos que já identificam o aumento na dosagem desse composto para a

aniquilação de ervas menos susceptíveis, como visto na Figura 2 (KOHLHASE; EDWARDS; OWEN, 2018; O'BRIEN; DAVIS; RIECHERS, 2018).

Figura 2 - Representação da dosagem aplicada de mesotriona em ervas daninha, cuja resistência genética foi desenvolvida.



Legenda: Demonstração do aumento da taxa de aplicação do herbicida Mesotriona à gerações de ervas daninha cuja evolução gerou espécies com maior quantidade de genes resistentes.

Fonte: Adaptado de Kohlhase, Edwards e Owen (2018).

Assim, conhecidas as características estruturais do herbicida que garantem sua capacidade em inibir a HPPD, diversos estudos a fim de propor novas moléculas que consigam contornar os efeitos de resistência são publicados (HE *et al.*, 2020; LI *et al.*, 2018), buscando não somente maior eficácia na mitigação de pragas, como também segurança à plantação através de sua seletividade, baixa dosagem e segurança aos consumidores, produtores e meio ambiente.

O emprego do método QSAR é promissor nos avanços para a obtenção de um novo composto eficaz, pois através deste, a proposição de potenciais candidatos é possível com a redução do emprego de análises exploratórias, que são geralmente caras e demoradas (DUARTE, 2016). Nessa perspectiva, as vantagens do método MIA-QSAR o torna essencial no estudo de novos compostos, trazendo resultados igualmente eficazes a menor custo computacional se comparado às técnicas análogas (DUARTE, 2016). A variação *augmented* do método traz ainda contribuições indispensáveis, tornando possível o tratamento de propriedades periódicas como raio de Van der Waals e eletronegatividade, além de seus efeitos na obtenção de um resultado plausível e aplicável.

Métodos de *docking* molecular trazem ainda resultados complementares sobre as interações do ligante no sítio ativo enzimático, sendo indispensáveis em pesquisas computacionais que visem à proposição de novas moléculas com potenciais biológicos por simular, através de cálculos que objetivam encontrar os valores mínimos de energia, as conformações mais prováveis das novas propostas na macromolécula de interesse.

Desse modo, dois conjuntos de moléculas foram submetidos às técnicas de MIA-QSAR e *docking* molecular, na tentativa de 1) reconhecer fatores que influenciam diretamente na atividade de uma molécula biologicamente ativa; 2) propor candidatos mais eficientes com base em pesquisas de novos herbicidas tricetônicos; 3) avaliar a performance das propostas mediante comparações entre as melhores moléculas do conjunto e o padrão mesotriona e 4) identificar o coeficiente de partição destes compostos, que traz grandes contribuições sobre efeitos de segurança ambiental e humana.

2 OBJETIVOS

2.1 *Objetivo geral*

Construir modelos MIA-QSAR das atividades herbicidas (pK_i e pIC_{50}) para dois conjuntos de dados congêneres, bem como compreender os aspectos químicos responsáveis pelas suas atividades biológicas, por meio de estudos de *docking* e mapas de contorno MIA.

2.2 *Objetivos específicos*

- 1- Elaborar imagens moleculares dos derivados de tricetonas usando o programa GaussView e convertê-las em matriz de descritores MIA;
- 2- Realizar regressão por mínimos quadrados parciais (PLS) dos descritores MIA com as correspondentes atividades herbicidas utilizando o programa Chemoface;
- 3- Validar os modelos MIA-QSAR construídos e utilizá-los para a previsão da atividade herbicida de novos candidatos a agroquímicos;
- 4- Construir MIA-Plots para interpretar quimicamente os modelos MIA-QSAR;
- 5- Realizar estudos de *docking* molecular para validar as possíveis previsões obtidas *via* MIA-QSAR e compreender o modo de interação entre as moléculas eventualmente propostas e o seu sítio de ação na enzima HPPD (EC 1.13.11.27, código PDB: 6M6D).

3 REFERENCIAL TEÓRICO

3.1 *Ervas daninhas e herbicidas*

Uma das grandes preocupações do manejo agrícola é o dano em lavouras causado pelo surgimento de ervas daninha que, por definição, são “quaisquer plantas de caráter silvestre que crescem em áreas controladas pelos humanos como cultivos, hortas ou jardins, e cuja presença é indesejada”¹. Essas agem de modo competitivo com a plantação sorvendo água, nutrientes, espaço e luz solar (HE *et al.*, 2020) e resultam em significativos prejuízos no que se refere à qualidade de produtos e rendimento da colheita, podendo gerar perdas que variam de 36% a 80% em casos onde não são controladas (KARAM *et al.*, 2018) e refletir em impactos nos setores econômicos, comerciais, agrícolas, entre outros.

Muitas ideias foram consolidadas no que se refere à aniquilação dessas plantas no campo de cultivo. Práticas mecânicas, térmicas, alelopáticas e biológicas, por exemplo, são reconhecidas como métodos alternativos para o controle das ervas daninhas (COSTA *et al.*, 2018), mas o uso de princípios químicos visando seu potencial herbicida apresenta grande aceitação dada a praticidade e facilidade de manejo², eficiência no combate seletivo às pragas, alta diversidade de produtos que atendem demandas específicas e economia em detrimento a outros métodos de controle (OLIVEIRA JR.; CONSTANTIN; INOUE, 2011).

Relatos sobre o emprego de produtos químicos para o controle de espécies indesejáveis no campo de cultivo remontam ao início do século XX com o uso de petróleo, sal e ácido sulfúrico (ROMAN *et al.*, 2007; ZIMDAHL, 2018a). Contudo, somente no ano de 1932 o produto de nome comercial *Dinoseb*, reconhecido como o primeiro herbicida orgânico sintético e seletivo, foi introduzido na França, dispondo de um mecanismo de ação atuante no desacoplamento da fosforilação oxidativa (ROMAN *et al.*, 2007; ZIMDAHL, 2018a) e servindo de incentivo ao desenvolvimento de novos herbicidas sintéticos e seletivos.

Desde então, diversos estudos contribuíram relevantemente para o desenvolvimento de novos princípios herbicidas (HUANG *et al.*, 2021; QU *et al.*, 2021), trazendo grandes avanços

¹ IBERDROLA. **Ervas daninha: quais são e porque não são tão terríveis.** Disponível em: <https://www.iberdrola.com/sustentabilidade/ervas-daninhas>. Acesso em: 19 dez. 2022.

² CROPLIFE BRASIL. **Herbicidas: seu uso favorece o desenvolvimento das plantas cultivadas.** Disponível em: <https://croplifebrasil.org/conceitos/herbicidas-seu-uso-favorece-o-desenvolvimento-das-plantas-cultivadas/>. Acesso em: 08 fev. 2023.

referentes ao conhecimento do mecanismo de ação em seu alvo biológico e classificações quanto a suas aplicabilidades, além de permitir que rigorosos processos fossem adotados para a certificação do potencial e empregabilidade no controle de espécies infestantes, com destaque aos testes *in vitro*, *in vivo* e *in silico*.

Os dados experimentais obtidos em testes *in vitro* trazem perspectivas interessantes nas potencialidades do herbicida em atuar, de modo controlado, na inibição de enzimas e na formação de compostos essenciais. Nesses testes, alguns parâmetros quantitativos mensurados, como por exemplo a concentração inibitória a 50% (IC₅₀) e a constante de inibição (K_i), servem de métrica à avaliação e defesa do desempenho desses compostos na cultura, relacionando grandezas matemáticas a propriedades biológicas de interesse (BURLINGHAM; WIDLANSKI, 2003).

O parâmetro IC₅₀, cujo cálculo de dá pela formula disposta no item 3.1.1, é responsável por quantificar, em um sistema competitivo, a concentração de inibidor necessária para que a taxa de uma reação enzimaticamente catalisada seja reduzida à metade se comparada à taxa de um mesmo sistema não inibido (BURLINGHAM; WIDLANSKI, 2003). Esse parâmetro, que se relaciona à potência do composto de modo inversamente proporcional, apresenta resultado menos preciso que constantes físico-químicas, por não descrever diretamente o equilíbrio químico da reação (BURLINGHAM; WIDLANSKI, 2003), sendo, porém, frequentemente empregado em ensaios de performances *in vitro*, por apresentar maior simplicidade e redução de suposições necessárias. Seu cálculo é realizado em função das constantes de inibição (K_i) e de Michaelis-Menten (K_M).

$$IC_{50} = K_i \left(1 + \frac{[S]}{K_M} \right) \quad (3.1.1)$$

A simplicidade da obtenção do IC₅₀ é confirmada na comparação entre os métodos citados. Enquanto para o cálculo de K_i há a elaboração de curvas com variações na concentração do substrato para cada concentração de inibidor de uma faixa delimitada, requerendo geralmente mais de 75 taxas para uma análise realizada em triplicata, o IC₅₀ pode ser determinado com apenas uma concentração de substrato em uma faixa variável de inibidor, reduzindo os dados para até 15% do que seria esperado no cálculo de constantes (BURLINGHAM; WIDLANSKI, 2003). Por outro lado, o tratamento fixo da concentração de substrato oferece ao IC₅₀ resultados relativos, podendo este valor apresentar diferenciações para concentrações diferentes da utilizada no desenvolvimento do experimento (BURLINGHAM; WIDLANSKI, 2003).

De modo complementar, as análises *in vivo* são essenciais no tratamento dinâmico da efetividade do herbicida, testando condições reais de aplicação e possíveis influências que afetam o composto em desvios da idealidade. Assim, conhecidas as características vegetais e os efeitos desejáveis no emprego de princípios químicos, o composto de interesse deve atender alguns requisitos para ser considerado bom herbicida. Aos herbicidas foliares, por exemplo, esperam-se que sejam capazes de serem retidos pela folhagem, penetrarem a cutícula, serem móveis ao fluido aquoso envolto à célula e de adentrarem-na através da membrana plasmática, a fim de atuar em sua região adequada (ROMAN *et al.*, 2007). Exigências quanto à seletividade são também impostas, assegurando que a metabolização do composto seja feita pela cultura, mas não seja realizada pela erva daninha, atuando assim para sua erradicação e garantindo segurança à plantação (CASIDA, 2017). Alguns parâmetros químicos são frequentemente adotados na dedução de possíveis comportamentos do herbicida no metabolismo vegetal, orientando pesquisas ao desenvolvimento de princípios compatíveis e otimizados, além de auxiliar na interpretação dos efeitos verificados.

O parâmetro químico que melhor define a mobilidade de determinado composto é o coeficiente de partição octanol/água ($\log P$), que correlaciona a estrutura química à lipofilicidade, com valores que trazem informações sobre a preferência do produto químico em interagir com fases polares ou apolares (BENFENATI *et al.*, 2003). Desse modo, ele auxilia na identificação de compostos que possam apresentar boa mobilidade aos fluidos vegetais e permeabilidade às membranas e cutícula, além de ser indicativo de possíveis regiões em que esses se depositam por conta de sua maior afinidade (EROS *et al.*, 2002). No transporte de herbicidas, esse parâmetro apresenta grande importância, visto que no xilema há a predominância de substâncias hidrofílicas e, no floema, as lipofílicas, sendo as de natureza lipofílica absorvidas com maior rapidez dadas as características dos componentes estruturais das folhas, caules e raízes (OLIVEIRA JR.; CONSTANTIN; INOUE, 2011).

Aspectos de segurança ambiental são também tratados no estudo de herbicidas. O coeficiente de partição, nesse escopo, relaciona-se diretamente à capacidade do herbicida ser adsorvido ao carbono orgânico do solo e não ser lixiviado por chuvas e irrigações, possibilitando aferir ainda sobre possíveis riscos, tais como a bioacumulação em organismos de vida aquática e na cadeia alimentar (BENFENATI *et al.*, 2003).

Dada a dificuldade encontrada na obtenção experimental do $\log P$, o uso de softwares na predição do comportamento lipofílico de moléculas é comumente observado, sendo o *Molinspiration* um bom recurso por apresentar fácil manuseio, gratuidade e robustez

conferida por validações estatísticas (PYKA; BABUŠKA; ZACHARIASZ, 2006). O algoritmo do site, calibrado através do pareamento de valores experimentais e calculados de um montante superior a 12 mil amostras, define o coeficiente de partição através da análise de contribuições de grupo, dispondo registro de 35 fragmentos básicos e 185 maiores, com considerações sobre efeitos de ligações de hidrogênio e interações de carga³.

Apesar das viabilidades e consolidação do uso de herbicidas no controle de pragas, a resistência, compreendida “como a habilidade da erva em resistir à exposição a um agente nocivo sem sofrer seus efeitos”(ZIMDAHL, 2018c), surge no contexto como grande desafio, sendo já descrito o fenômeno de resistência em vinte e três mecanismos de ação em um total de vinte e seis conhecidos (KARAM *et al.*, 2018). Diferente da tolerância, que pode ser entendida como a “habilidade da erva em desempenhar suas funções e crescimento após exposta a agentes nocivos à outras espécies” (ZIMDAHL, 2018c), o que traz benefícios ao desenvolvimento de herbicidas no que se refere à seletividade de espécies, a resistência, por se tratar de um fenômeno biológico evolutivo, gera preocupação, pois tende a reduzir a susceptibilidade de ervas ao composto até então empregado.

No mecanismo de inibição da enzima 4-hidroxifenilpiruvato dioxigenase (HPPD), por exemplo, o problema de resistência à mesotriona já foi observado para as espécies *Arabidopsis thaliana* (TUKUR; SHALLANGWA; IBRAHIM, 2019), *Amaranthus tuberculatus* e *Palmer amaranth*, trazendo más perspectivas para a efetividade do herbicida mais empregado da classe das tricetonas. As plantas da espécie *amaranth* são ervas que, se não controladas, podem ser responsáveis por reduções que variam de 56% a 74% na colheita de milho e soja, sendo já encontradas práticas de combinação de atrazina à mesotriona, a fim de melhorar sua absorção e translocação (CHAHAL; JUGULAM; JHALA, 2019; JACOBS *et al.*, 2020). Esse fenômeno pode estar relacionado à catalise realizada pelo citocromo P450, que age em uma rápida metabolização oxidativa do herbicida original (JACOBS *et al.*, 2020), impedindo que este chegue íntegro ao sítio receptor e realize sua função.

Manejos inadequados como o aumento da dosagem do herbicida e consequentes prejuízos ao meio ambiente são geralmente resultantes desse problema, motivando investigações no incremento de práticas pré e pós emergentes, com sugestões que vão desde potencialização por combinações químicas, rotações de cultura e de herbicidas de atuações diferentes, etc. (CHAHAL; JUGULAM; JHALA, 2019), a pesquisas que visam modificações

³ MOLINSPIRATION. **logP** – **octanol-water partition coefficient**. Disponível em: <https://www.molinspiration.com/services/logp.html>. Acesso em: 22 dez. 2022.

tanto genéticas à cultura quanto estruturais aos compostos até então utilizados (CASIDA, 2017).

Frente aos esforços cedidos para a mitigação desse problema, torna-se evidente a importância da busca por melhorias na superação desses desafios, com ênfase no que se refere ao extermínio das ocorridas combinações químicas ausentes nos relatórios de autorização que resultam em sistemas complexos de efeitos toxicológicos e ambientais desconhecidos (DUMAS *et al.*, 2017). Visam-se, ainda, com novas pesquisas, benefícios como melhorias na seletividade e produtos de maior eficácia e segurança ao meio ambiente e à humanidade, propondo candidatos que apresentem bom equilíbrio entre potência, mobilidade e estabilidade (HUANG *et al.*, 2021).

3.2 HPPD

A enzima HPPD (Figura 3), presente em todos os seres vivos aeróbicos com a exceção de algumas bactérias gram-positivas (MORAN, 2005), atua na rota metabólica da L-tirosina no segmento da tirosina aminotransferase, fornecendo acetoacetato e fumarato como fontes energéticas, como ilustrado na Figura 4, além de outros produtos essenciais na manutenção e funcionamento de organismos mais complexos (MORAN, 2005).

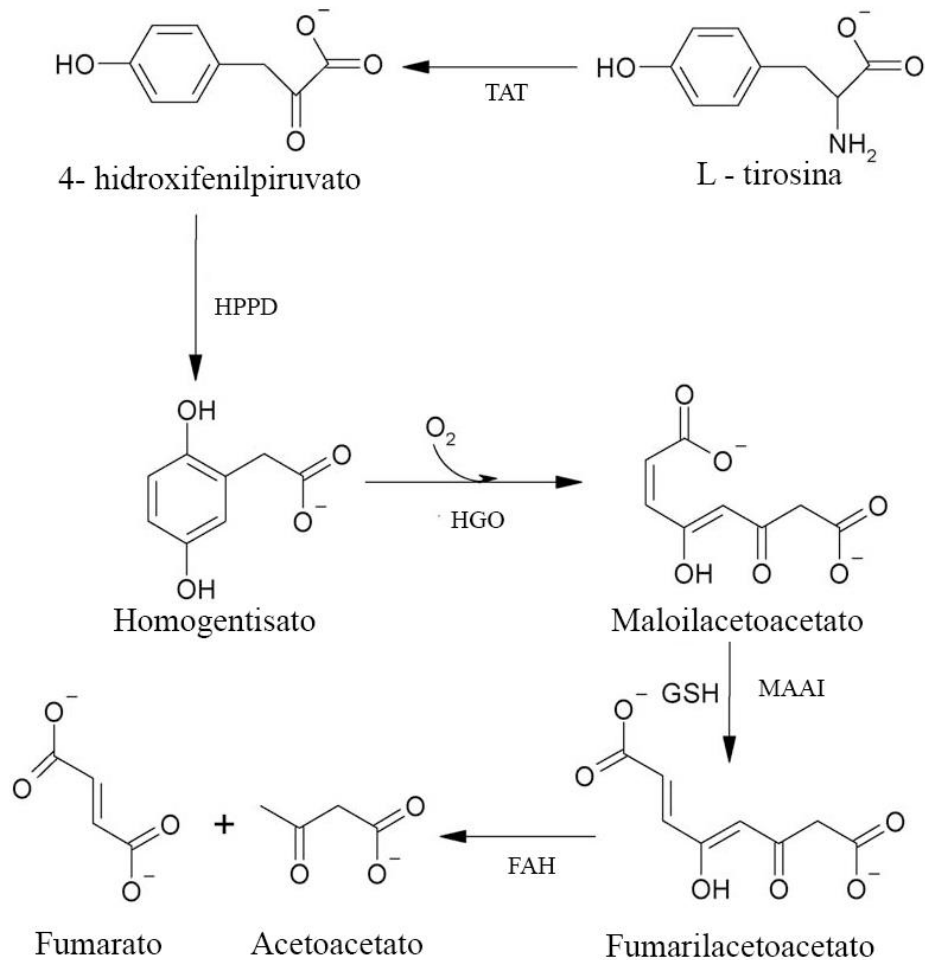
Figura 3 – Representação tridimensional da estrutura da enzima HPPD.



Fonte: RCSB PDB (6M6D)⁴.

⁴ RCSB PDB. **6M6D**. Disponível em: <https://www.rcsb.org/structure/6M6D>. Acesso em: 22 fev. 23.

Figura 4 – Rota metabólica vegetal da *L*-tirosina pelo seguimento da enzima TAT.

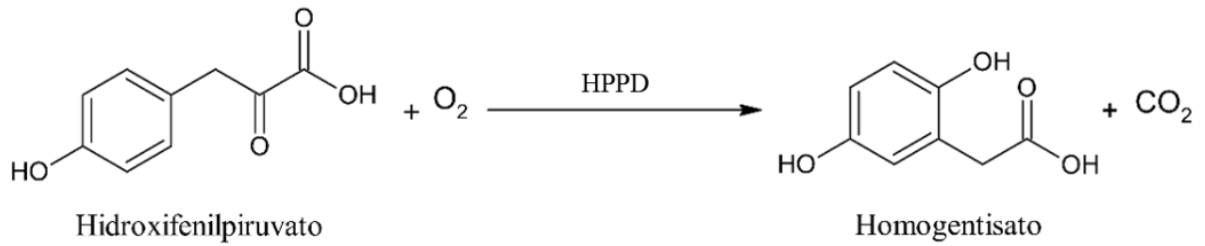


Legenda: TAT – Tirosina Aminotransferase; HPPD – Hidroxifenilpiruvato Dioxigenase; HGO – Homogentisato; MAAI – Maloilacetoacetato Isomerase; FAH – Fumarilacetoacetato hidrolase.

Fonte: Adaptado de Dixon e Edwards (2006).

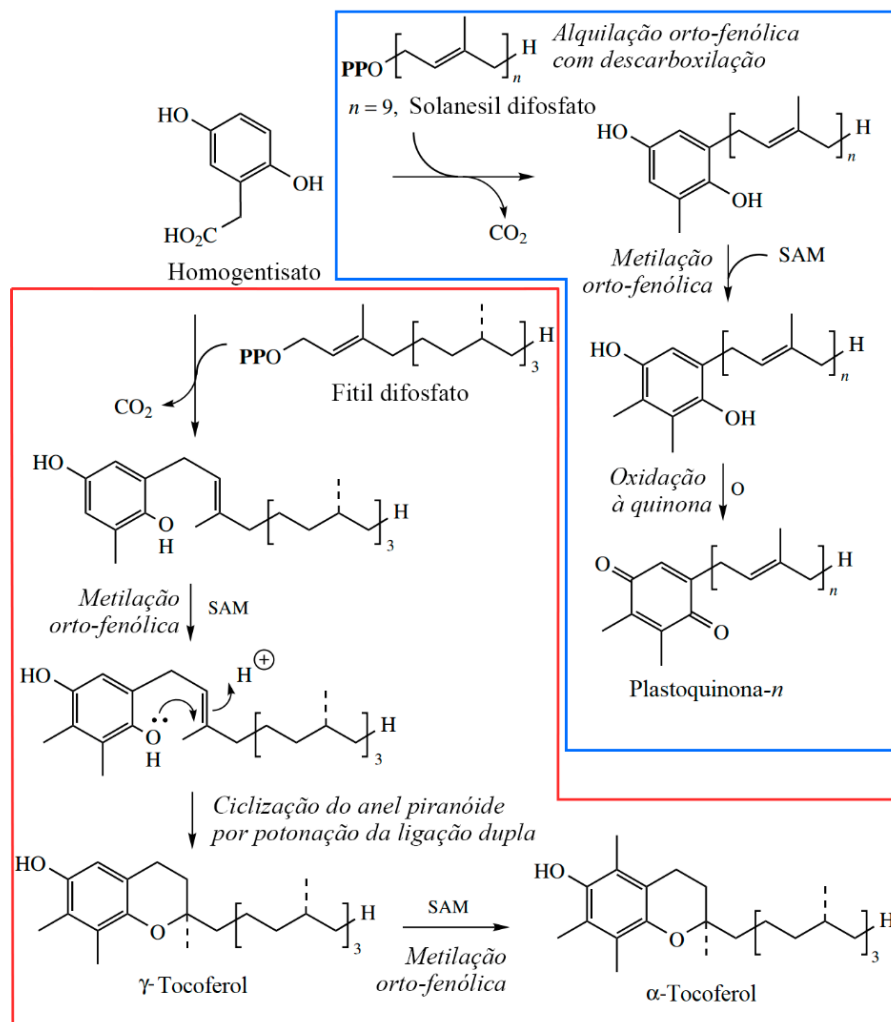
No metabolismo vegetal, esta enzima apresenta grande importância para o crescimento e evolução da planta, convertendo hidroxifenilpiruvato a homogentisato através da reação demonstrada na Figura 5 (TUKUR; SHALLANGWA; IBRAHIM, 2019). O homogentisato, também conhecido como ácido homogentísico, participa como precursor na síntese dos cofatores essenciais plastoquinona e α -tocoferol (XU *et al.*, 2020), sendo transformado através das reações que ocorrem nas rotas demonstradas na Figura 6. Esses cofatores permitem o correto funcionamento da fotossíntese, são essenciais na síntese de carotenóides e desempenham, como principal função, a inibição de processos foto-oxidativos por espécies reativas altamente prejudiciais à membrana tilacóide (KUPPER *et al.*, 2018).

Figura 5: Conversão do hidroxifenilpiruvato à homogentisato por ação da HPPD



Fonte: Do autor (2022).

Figura 6: Reações de formação dos cofatores α -Tocoferol e Plastoquinona.



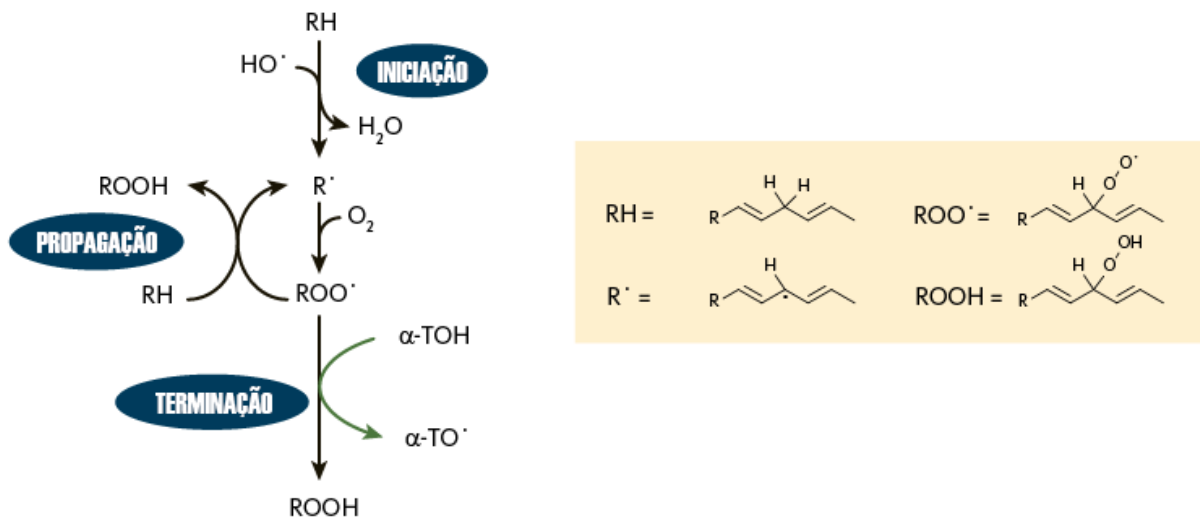
Legenda: Em destaque vermelho – rota metabólica de produção do α -Tocoferol. Em destaque azul – rota metabólica de produção da Plastoquinona. SAM – S-adenosil metiona.

Fonte: Adaptado de Wikipedia⁵

⁵ WIKIPEDIA. **Ruta del ácido shikímico**, 2022. Disponível em: https://es.wikipedia.org/wiki/Ruta_del_%C3%A1cido_shik%C3%ADmico. Acesso em: 22 fev. 2023.

Formado por derivados do metabolismo de aminoácidos aromáticos e de fitil-difosfato (DELLAPENNA; POGSON, 2006), o α -tocoferol compõe o conjunto de agentes antioxidantes de grande valor aos vegetais, atuando na interrupção da reação em cadeia da peroxidação lipídica através de seu domínio redutor fenólico (GALLI, 2012). Em sua atuação química, os tococromanóis, classe em que se encontram os tocoferóis e tocotrienóis, interagem com os ácidos graxos poli-insaturados da membrana a fim de proteger o dano oxidativo, reagindo de forma radicalar na oferta de um radical hidrogênio para a neutralização dos radicais peroxilipídicos e oxigênios reativos (Figura 7) e, posteriormente, reciclado ou convertido a quinona (DELLAPENNA; POGSON, 2006). O radical alfa-tocoferila pode ser ainda um dos produtos encontrados após o processo de terminação da reação em cadeia, sendo então reduzido por ação do ácido ascórbico (CERQUEIRA; MEDEIROS; AUGUSTO, 2007), como ilustrado na figura 8.

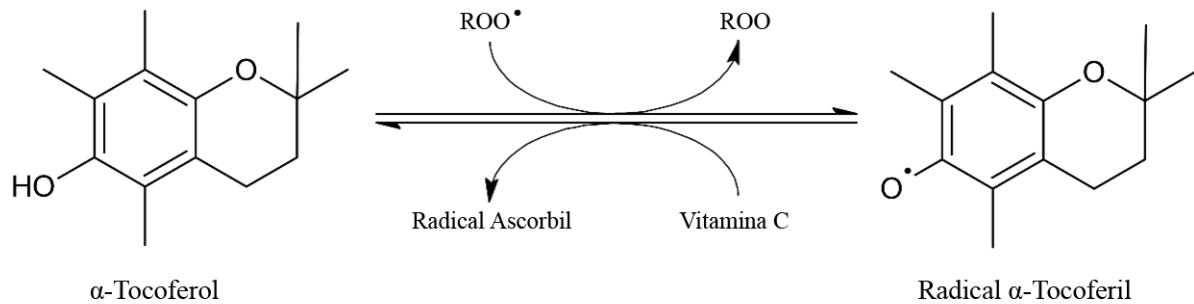
Figura 7: Reações de iniciação, propagação e terminação de radicais envolvendo lipídios e α -Tocoferol.



Fonte: Adaptado de Linus Pauling Institute (2018).⁶

⁶ OREGON STATE UNIVERSITY. **Coenzyme Q10**. Disponível em: <https://lpi.oregonstate.edu/mic/dietary-factors/coenzyme-Q10>. Acesso em: 22 fev. 2023.

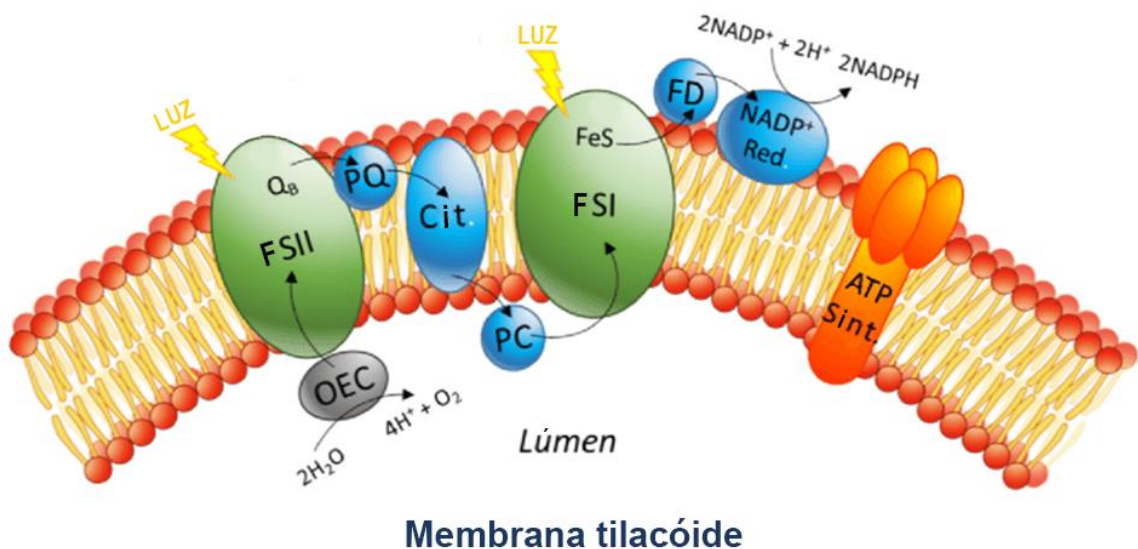
Figura 8: Reação de neutralização de radical livre em lipídeos e reciclagem do α -Tocoferol por sacrifício do ácido ascórbico (Vitamina C).



Fonte: Adaptado de Guaratini, Medeiros e Colepicolo (2007).

A plastoquinona, também formada na rota metabólica da *L*-tirosina, apresenta-se no processo de fotossíntese da planta atuando na recepção dos elétrons da clorofila P680 luminosamente excitada no fotossistema II e os transportando ao citocromo b6f, garantindo que a cadeia transportadora de elétrons funcione de modo contínuo na produção de energia fotossintética (NORRIS; BARRETTE; DELLAPENNA, 1995), como visto na Figura 9.

Figura 9 - Representação esquemática da cadeia transportadora fotossintética na membrana tilacóide.

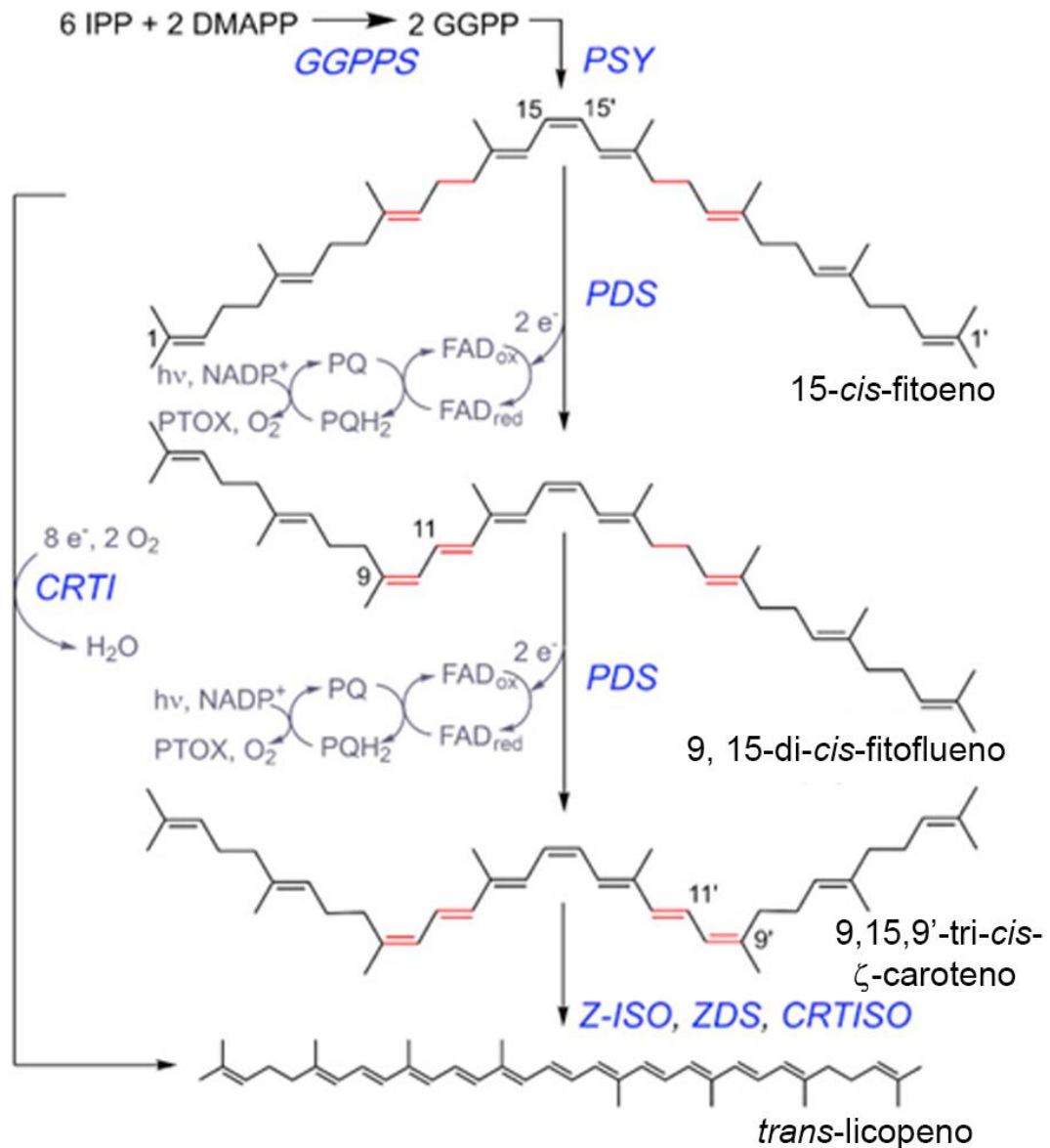


Legenda: FSII – fotossistema II; FSI – fotossistema I; Cit – citocromo; OEC – complexo envolvendo oxigênio; PC – plastocianina; PQ -plastoquinona; FD – ferredoxina.

Fonte: Adaptado de Tucci *et al.* (2019).

Ela também desempenha importante papel na produção dos pigmentos carotenoides, atuando como cofator na recepção dos elétrons gerados na conversão de fitoeno a fitoflueno, que são precursores saturados dos compostos carotenoides (BRAUSEMANN *et al.*, 2017). Os elétrons por ela transportados são oriundos das ocorrências de reação de eliminação, recebidos anteriormente pelo cofator FAD, como pode ser visto no esquema da Figura 10.

Figura 10 - Dessaturação do caroteno em plantas.



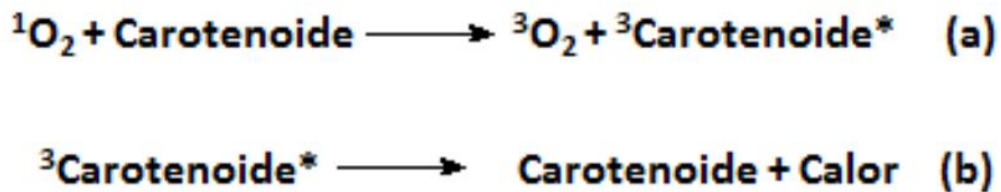
Fonte: Adaptado de Brausemann *et al.*, 2017.

Os compostos corantes dispostos na membrana tilacóide desempenham importante papel na proteção e produção de energia através da incidência de luz solar. As clorofilas, por exemplo, refletem coloração verde por absorver radiação luminosa nos comprimentos

correspondentes às cores complementares, sendo a radiação por elas absorvidas convertidas a energia química para a ocorrência do processo de fotossíntese.

Juntamente à clorofila, os carotenoides, pigmentos lipossolúveis de grande cadeia insaturada, são dispostos em proporções diversas nas diferentes espécies vegetais e auxiliam na regulação da quantidade de luz incidida nas organelas responsáveis pela produção de energia, absorvendo a radiação em comprimentos que diferem das colorações que vão de amarelo a vermelho (NORRIS; BARRETTE; DELLAPENNA, 1995). Esses compostos são também responsáveis por garantir às plantas suas cores características, funcionando de modo eficaz na proteção à oxidação através da dissipação do elétron de oxigênio singlete em forma de calor, como visto na Figura 11. Os tocoferóis e a plastoquinona, assim como o caroteno, são também agentes antioxidantes que auxiliam na manutenção de radicais livres nas plantas, inibindo o seu stress oxidativo e a degradação dos compostos carotenoides (KUPPER *et al.*, 2018)

Figura 11 – Reação de supressão do oxigênio singlete por ação de carotenoide.



Legenda: (a) reação de conversão do oxigênio singlete em oxigênio tripleto; (b) reação subsequente, em que o carotenoide tripleto retorna ao estado fundamental, liberando o excesso de energia sob a forma de calor.

Fonte: MESQUITA; TEIXEIRA; SERVULO (2017).

Na ausência de homogentisato, condição ocasionada pela inibição da enzima HPPD, as moléculas de tocoferol e plastoquinona deixam de ser produzidas, resultando nas indevidas manutenções dos compostos antioxidantes, do fluxo de elétrons do fotossistema II e da proteção a alta incidência de luz solar. Assim, dada a indisponibilidade de controladores do fluxo de elétrons, a luz experimentada na ausência das proteções pigmentares resulta em estímulo nos componentes fotossintéticos que desencadeiam a produção de agentes reativos que são liberados no meio celular. Esses agentes, em concentrações elevadas, atacam os compostos reguladores remanescentes, potencializando o processo depreciativo vegetal. Os herbicidas que visam à inibição da HPPD e, conseqüentemente, à manutenção de compostos

pigmentares, são conhecidos como herbicidas descolorantes, pois agem para que a planta apresente branqueamento, murchamento, necrose e seca pelo efeito do agente utilizado (MORAN, 2005; TUKUR; SHALLANGWA; IBRAHIM, 2019).

O processo de descoloração das folhagens ocorre devido à clorofila, com a perda de seus agentes de foto proteção, apresentar-se de modo vulnerável à exposição a luz solar, sendo fotodegradada em um processo oxidativo (O'BRIEN; DAVIS; RIECHERS, 2018), seguido por danos à membrana do cloroplasto e aos tilacóides, dada a formação de clorofila tripleto e oxigênio singleto (O'BRIEN; DAVIS; RIECHERS, 2018). Essa espécie de oxigênio extremamente reativa comporta-se como um radical livre e catalisa o processo degradativo dos carotenoides remanescentes com a ausência dos tocoferóis e plastoquinona (KUPPER *et al.*, 2018), mantendo a planta em um estado de decomposição severo. Isso resulta na paralisação do processo de fotossíntese (TUKUR; SHALLANGWA; IBRAHIM, 2019), que apresenta grande importância vegetal, dada a conversão de gás carbônico a glicose, produto precursor na formação de outros compostos orgânicos essenciais ao desenvolvimento da planta.

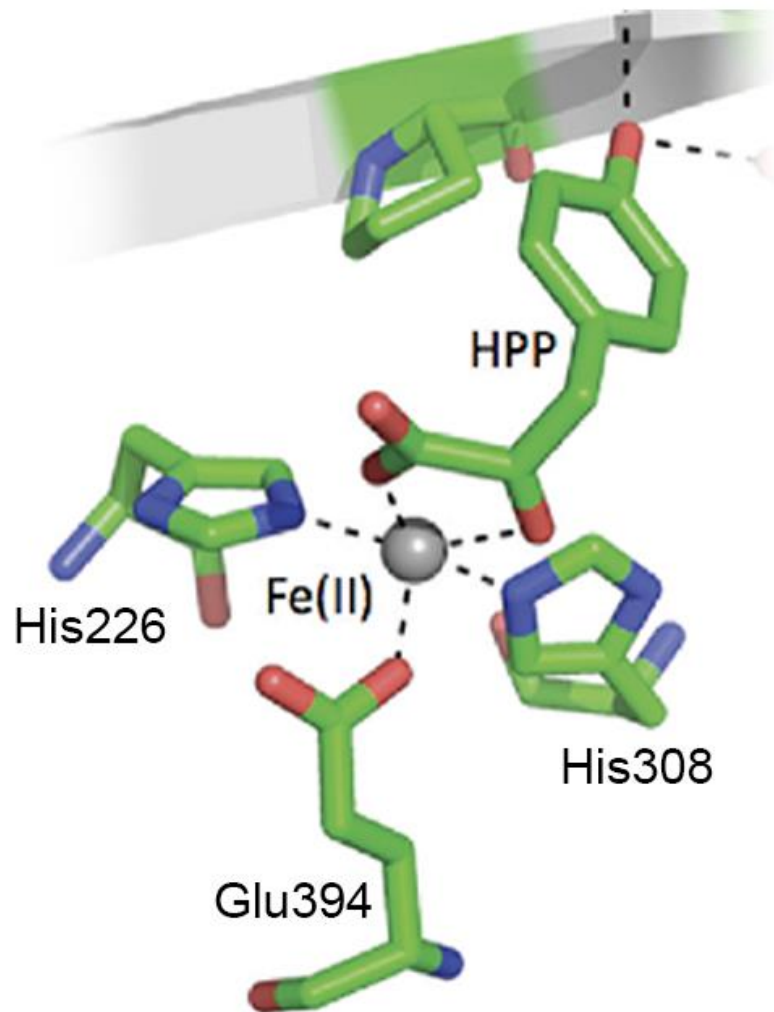
A enzima HPPD pode ser inativada por compostos da classe das tricetonas, isoxazóis e pirazolonas (JACOBS *et al.*, 2020). Até o ano de 2015, já comercializavam 13 compostos das referidas classes (WANG *et al.*, 2015), com notória utilização dos agentes mesotriona, biciclopirona, tembotriona, topamezona e isoxaflotol nas lavouras de milho, pirazolato e pirazoxifeno em culturas de arroz e pirasulfatol no controle de ervas em trigo (HAWKES *et al.*, 2019).

O emprego de herbicidas inibidores da enzima HPPD é orientado, geralmente, a plantas de folhas largas, sendo vantajoso por normalmente apresentarem baixas taxas de aplicação e toxicidade, seletividade em amplo espectro de ervas daninhas e boa segurança ambiental (FU *et al.*, 2021). Pelo reconhecimento do modo de ação desses herbicidas, suas potencialidades farmacológicas também são exploradas para tratamento da doença tirosinemia tipo I através da inibição da HPPD em humanos (MOLCHANOV; GRYFF-KELLER, 2009).

Por conta da eficiência reduzida dos atuais herbicidas utilizados no controle das ervas daninhas pelo método de descoloração, pesquisas são ainda realizadas objetivando a descoberta de novos compostos que apresentem melhores potencialidades direcionadas à capacidade agroquímica, sendo imprescindível o conhecimento do mecanismo e atuação da enzima para que novas possibilidades possam ser avaliadas (ZIMDAHL, 2018b).

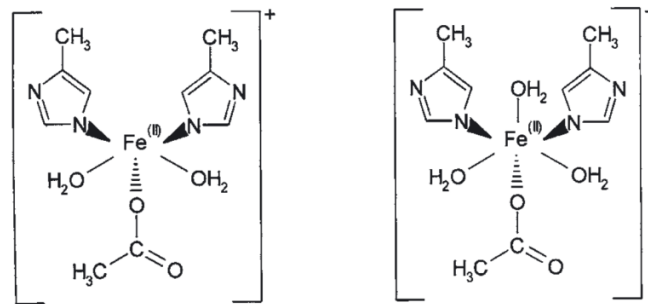
Em termos estruturais, o sítio ativo da HPPD, como visto na figura 12, é funcional na presença de um íon Fe (II), que se coordena a duas histidinas e um glutamato dispostos na cadeia polipeptídica da enzima (WANG *et al.*, 2021). O íon ferroso pode ser encontrado em suas formas pentavalente e hexavalente, permanecendo em uma mistura em equilíbrio dinâmico, sendo as outras ligações coordenadas formadas pelo efeito de hidratação (MOLCHANOV; GRYFF-KELLER, 2009), como visto na Figura 13. No complexo enzima-substrato, as ligações até então aquosas são desfeitas e direcionadas a interagir com a molécula de hidroxifenilpiruvato (MOLCHANOV; GRYFF-KELLER, 2009), convertendo o substrato a homogentisato através do mecanismo de reação expresso na Figura 14.

Figura 12: Representação do sítio ativo da enzima HPPD com o ligante HPP ancorado ao átomo metálico de Fe II.



Fonte: Adaptado de Shah *et al.* (2011).

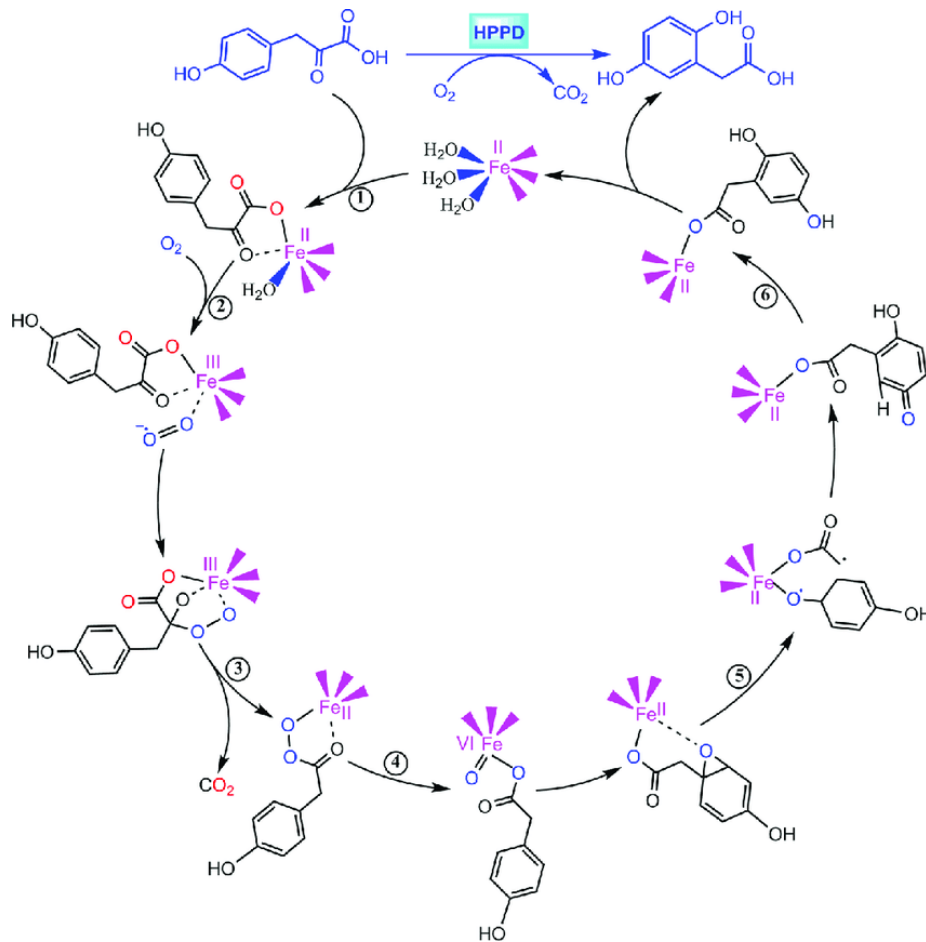
Figura 13 - Complexo ferroso nas formas pentavalente e hexavalente no sítio ativo da enzima HPPD.



Legenda: A presença de uma molécula de acetato no local correspondente ao glutamato é justificada pela cristalização da enzima em tampão acetato (MOLCHANOV; GRYFF-KELLER, 2009).

Fonte: MOLCHANOV; GRYFF-KELLER (2009).

Figura 14 - Mecanismo simplificado da catálise da HPPD.



Legenda: (1) formação do complexo HPPD-Fe(II); (2) adição dioxigenada; (3) descarboxilação; (4) clivagem do O-O; (5) migração da cadeia de ácido acético; (6) hidroxilação em C1.

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

Alguns modelos tridimensionais da HPPD apresentam no sítio ativo enzimático o íon cobalto II em substituição ao íon ferroso. A observação ocorre em casos severos cujas condições de cristalização da enzima exigem a presença de cloreto de cobalto (SANTUCCI *et al.*, 2017), sendo já possível encontrar trabalhos *in silico* de *docking* molecular com o emprego desta variável (QU *et al.*, 2021).

Desse modo, os inibidores da HPPD buscam interagir de forma coordenada com o íon ferroso presente no sítio ativo enzimático deslocando as moléculas de água que com ele interagem a fim de impedir, de modo competitivo, que o complexo enzima/substrato seja formado.

3.3 *Tricetonas*

Em meados da década de 70, ao observar o crescimento reduzido de algumas ervas nas proximidades da árvore Malaleuca-Citrina, Rocaboy-Faquet questionou a influência da árvore sobre suas observações (FU *et al.*, 2019a), chegando à conclusão de que o fenômeno observado se tratava de efeitos alelopáticos dada a produção do composto leptospermona, conhecido desde 1921 através de estudos de Penfold sobre produtos naturais (CORNES, 2005).

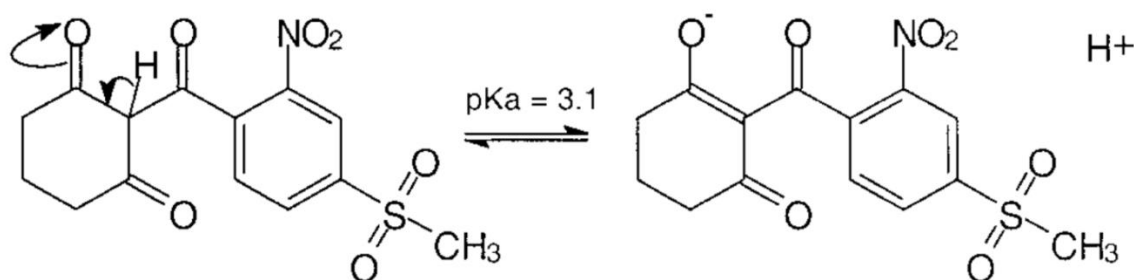
Na tentativa de expandir os efeitos da leptospermona à escala industrial, a mesotriona foi então projetada à semelhança do produto alelopático e sintetizada pela empresa Syngenta sob nome comercial *Callisto*, apresentando grande aceitabilidade, dada a alta seletividade, eficácia de controle e alta segurança em termos de toxicidade (CORNES, 2005; (BEAUDEGNIES *et al.*, 2009).

Apesar de suas potencialidades herbicidas, o fenômeno de resistência tende a torná-lo ineficiente no controle de pragas. A instabilidade da mesotriona durante o transporte, armazenamento e aplicação também orientam cada vez mais estudos na perspectiva de seu melhoramento (WANG *et al.*, 2019), com pesquisas que visam a observação das relações de subprodutos com o solo no que se refere ao pH e adsorção (DYSON, 2002).

Os ligantes da classe das β -tricetonas, por possuírem carbonilas em posições estratégicas, mostram-se promissores na interação com o átomo metálico da enzima HPPD. Estudos de *docking* evidenciam interações desses compostos de modo semelhante ao hidroxifenilpiruvato, com a efetivação de duas ligações coordenadas oriundas de pares de

elétrons livres dos oxigênios carbonílicos. O fenômeno de tautomerização, conhecido na área de química como a condição em que dois isômeros coexistem em equilíbrio, passou a ser considerado com os avanços dos estudos em tricetonas. Já se sabe o pK_a próximo a 3,1 para o hidrogênio α tri-carbonílico (DYSON, 2002), o que faz com que a estrutura mais representada em trabalhos *in silico* no tratamento de sistemas biológicos seja o tautômero enólico, como visto na Figura 15.

Figura 15 – Reação de enolização do composto Mesotriona.



Fonte: Adaptado de Dyson (2002).

Conhecidas as características representativas e de interesse na estrutura de um potencial inibidor, o método QSAR torna-se, então, uma técnica valiosa na pesquisa de novos produtos inibidores da HPPD que sejam eficientes e atendam aos requisitos propostos que os considerem um bom herbicida. Para isso, bases de dados que apresentam informações coletadas experimentalmente são úteis na verificação da potencialidade de alguns compostos, orientando os estudos rumo à identificação de fatores que fazem com que determinado produto apresente melhores resultados.

Alguns projetos relacionados à verificação do desempenho de classes de compostos têm sido desenvolvidos neste segmento, com estudos dirigidos a moléculas que apresentam três agrupamentos cetona distanciados em posições beta (DAYAN *et al.*, 2009), a quintriona, que apresentou bom desempenho e seletividade no emprego em cultura de arroz (PENG *et al.*, 2021), e sínteses de compostos tricetônicos de quinolina (WANG *et al.*, 2015) e cetonas aromáticas de pirazol (FU *et al.*, 2020), todos com peculiaridades que os tornam potenciais alternativos aos tradicionais inibidores da HPPD.

Outros também permitiram a descoberta de novos inibidores da HPPD híbridos, sendo descritos trabalhos com moléculas do tipo aroildicetônicos e aroiltricetônicos (FU *et al.*,

2019b), que se coordenam ao sítio ativo da enzima por um grupo carbonílico e hidroxílico; pirazol-quinazolina e análogos tricetônicos da benzoxazinona, que apresentaram candidatos com valor de IC_{50} aproximadamente 2,7 e 1,5 vezes mais potentes que o correspondente ao da mesotriona, respectivamente (HE *et al.*, 2020; LI *et al.*, 2018); derivados cetônicos de niacina e aril-formil piperidinona na utilização de precursores de origem natural para a síntese dos compostos herbicidas (FU *et al.*, 2021; FU *et al.*, 2019); e estudos feitos na modelagem QSAR que apresentaram bons resultados estatísticos no domínio dos híbridos de dicetona quinazolina (FU *et al.*, 2017b).

Dentre os métodos empregados nas análises relacionadas a potenciais alternativos a inibidores da HPPD, é perceptível a dominância de técnicas de modelagem molecular, sendo o mais frequente o *docking* molecular, que contribui com avaliações valiosas das interações ocorrentes entre os componentes do complexo enzimático investigado. É de grande importância ressaltar que, dentre os trabalhos encontrados, esse método é frequentemente acompanhado pelo QSAR, sobretudo as técnicas CoMFA, CoMSIA e métodos bidimensionais (FU *et al.*, 2019d). Também foi notada a presença de alguns estudos por *virtual screening* (FU *et al.*, 2019c; FU *et al.*, 2017a) e dinâmica molecular (FU *et al.*, 2019d), com todos os trabalhos acima descritos apresentando resultados promissores, que servem como embasamento e estímulo à investigação de novos inibidores atuantes na inibição da HPPD.

3.4 QSAR

Diversos métodos e recursos podem ser encontrados no que se refere à descoberta e desenvolvimento de novos compostos com atividades desejáveis como fármacos e herbicidas (QU *et al.*, 2021). Um exemplo é o software HerbPAD, que visa estabelecer e quantificar potencialidades desconhecidas de moléculas através de suas similaridades com herbicidas de atividades já definidas (HUANG *et al.*, 2021). De modo análogo, alguns métodos de mecânica molecular permitem uma eficaz predição do comportamento de determinado composto a partir de uma classe bem definida e delimitada, apresentando resultados capazes de ofertar interessantes candidatos finais para análises convencionais *in vitro* e *in vivo*, que são geralmente caras e demoradas (SETH; OJHA; ROY, 2020).

Um dos métodos computacionais responsável por essa predição é o QSAR (*Quantitative Structure-Activity Relationships*), que foi desenvolvido em 1964 por Hansch e Fujita na tentativa de prever respostas biológicas através de componentes estruturais de compostos químicos (HANSCH; FUJITA, 1964; VEERASAMY *et al.*, 2011). Esse método, em essência, busca analisar de modo quantitativo as relações entre a atividade bioquímica de moléculas com suas especificidades estruturais, sendo frequentemente empregado em pesquisas de novos compostos com potenciais farmacológicos, agroquímicos e ambientais.

O QSAR consiste na elaboração de uma equação plausível e validada estatisticamente, a qual permite estabelecer a relação de descritores moleculares com o valor correspondente a uma determinada propriedade de interesse, tornando possível descrever sistemas compostos por estruturas moleculares de mesmo domínio químico que possuam comportamento até então desconhecido (ROY; AMBURE; AHER, 2017). A equação, usualmente denominada de **modelo**, é compreendida pelo item 3.4.1, em que P corresponde a propriedades ambientais, biológicas ou físico-químicas obtidas experimentalmente ou calculadas, D , aos descritores utilizados na elaboração do modelo e X , às contribuições individuais de cada descritor na obtenção do resultado final (ROY; AMBURE; AHER, 2017).

$$P = X_0 + X_1D_1 + X_2D_2 + \dots + X_nD_n \quad (3.4.1)$$

Os descritores moleculares são o “resultado final de um procedimento lógico e matemático que transforma informação codificada dentro de uma representação simbólica de uma molécula em um número útil, ou o resultado de algum experimento padronizado” (CONSONNI; TODESCHINI, 2010). Em outras palavras, descritores podem ser compreendidos como valores numéricos que informam computacionalmente características estruturais e experimentais de determinada molécula, permitindo que a avaliação entre suas possíveis relações seja feita. A depender de sua natureza, eles são capazes de descrever fatores constitucionais, eletrostáticos, geométricos, topológicos e mecânico-quânticos das estruturas ao sistema, além de carregarem também informações quanto à sua dimensionalidade (ALVES *et al.*, 2018).

As informações dimensionais representadas por descritores são originadas no preparo e análise de dados do método QSAR em níveis que vão do 0D ao 7D. É perceptível, porém, a dominância de trabalhos que visem a utilização de descritores 2D e 3D, pela facilidade em serem computados, compreendidos e interpretados (SETH; OJHA; ROY, 2020). Descritores

obtidos em versões de dimensões superiores são de difícil manuseio, trazem maior complexidade e alta demanda computacional (FREITAS; RAMALHO, 2013).

Dentre as metodologias mais empregadas, os métodos tridimensionais CoMFA (CRAMER; PATTERSON; BUNCE, 1988) e CoMSIA (KLEBE; ABRAHAM; MIETZNER, 1994) destacam-se pelo grande número de publicações com análises de sucesso (CORMANICH; NUNES; FREITAS, 2012). Esses métodos, porém, apresentam ainda grande complexidade, dada a necessidade de triagens conformacionais e alinhamentos tridimensionais do conjunto de moléculas, tornando cada vez mais árduo seu uso para conjuntos com quantidades elevadas de amostras. Desse modo, partindo da premissa de que a imagem de uma estrutura química é capaz de ceder informações estruturais relevantes e trazer contribuições significativas no que se refere a suas propriedades e comportamentos intrínsecos (FREITAS; BROWN; MARTINS, 2005), o método bidimensional *multivariate image analysis* in QSAR (MIA-QSAR) foi desenvolvido, facilitando a análise de conjuntos relativamente grandes e interpretação dos resultados obtidos.

Aplicável a séries de amostras congêneres com resultados preditivos a baixos custos computacionais, rapidez e simplicidade, o MIA-QSARt foi o primeiro método QSAR criado na perspectiva da utilização de representações pictóricas de estruturas químicas como instrumentos fundamentais de análise (FREITAS; BROWN; MARTINS, 2005). O método foi desenvolvido visando a utilização de imagens da estrutura de moléculas em preto e branco, distinguindo os pixels RGB pretos, de valor bit igual a 0 dada a ausência de contribuições individuais das cores vermelho, azul e verde, e brancos, de valor igual a 765, correspondente à soma das contribuições máximas de 255 de cada canal.

Esse método atua no mapeamento dos píxeis de determinadas estruturas em valores binários, vista a utilização de apenas os valores 0 e 765 para definir os bits presentes no sistema. O bit corresponde ao valor que resulta da soma das contribuições individuais da intensidade de cada cor (RGB) que compõe o píxel. Uma matriz de descritores (valores atribuídos aos píxeis de determinada estrutura [BIT]) é então obtida em uma análise de várias imagens, sendo a quantidade de linhas da matriz equivalente à quantidade de imagens analisadas e os valores das colunas correspondentes ao valor do píxel em determinada coordenada.

É imprescindível que alguns cuidados sejam tomados na elaboração das imagens, visto essa ser a responsável pelo fornecimento da sequência de píxeis que será posteriormente

relacionada à característica atribuída. Desse modo, é de grande importância o desenho correto das estruturas químicas a serem utilizadas no modelo, mantendo as mesmas disposições dos átomos, dos grupos substituintes e direção da cadeia carbônica, além de utilizar apenas uma denominação para cada tipo de substituinte.

Com os avanços do método, novas representações estruturais foram consideradas, originando o aug-MIA-QSAR, análogo ao MIA-QSAR tradicional, que é distinguido por alguns melhoramentos onde o método inicial apresentava limitações. Suas vantagens em relação à primeira versão estão tanto na possibilidade do emprego de imagens de estruturas químicas com a representação de átomos em tamanhos proporcionais ao raio atômico de van der Waals, quanto na utilização de valores na escala RGB que varia de 0 a 765 para compor a natureza diferenciada dos átomos (FREITAS; BARIGYE; FREITAS, 2015). Por permitir o emprego de cores diversas na confecção das imagens das estruturas, os dados podem ser tratados em função de propriedades atômicas, tais como o raio de van der Waals e a eletronegatividade, visto que os valores de bit podem ser facilmente adaptados com auxílio de softwares computacionais, permitindo reconhecer a melhor dependência das propriedades moleculares em relação às propriedades periódicas de seus elementos constituintes. Assim, os descritores utilizados na análise deixam de apenas diferenciar os átomos presentes e passam a considerar e inserir informações sobre suas características que, no tratamento de sistemas biológicos, apresentam grande relevância no estudo de possíveis interações.

A aplicação do método aug-MIA-QSAR inicia-se na busca de conjuntos de moléculas de um mesmo domínio químico com certo grau de identidade estrutural que possuam, em uma base de dados, valores de atividades biológicas calculadas ou obtidas experimentalmente. A capacidade de os dados adquiridos serem combinados é avaliada pela semelhança dos procedimentos empregados para a obtenção dos valores de atividade e na escolha do composto padrão que serve de referência aos valores das outras moléculas do conjunto.

No momento da projeção das estruturas em software de desenho molecular, cálculos de otimização podem ser empregados no objetivo de tratar com maior fidelidade a conformação espacial da região congênica, a fim de garantir maior acurácia na análise das imagens. Os métodos semiempíricos são, nessa etapa, geralmente empregados por superarem o alto custo computacional de métodos *ab initio*, trazendo aproximações de integrais que permitem resultados suficientes quando contempladas as parametrizações. O método Austin Model 1 (AM1) baseado no método MNDO é muito utilizado em cálculos de mecânica

molecular dada a parametrização eficiente de cadeias carbônicas com heteroátomos, tais como oxigênio, nitrogênio e hidrogênio (DEWAR *et al.*, 1985).

A região congênica otimizada é, então, utilizada na obtenção das imagens referentes a cada amostra do conjunto, cuidando para que regiões variáveis na molécula não sejam omitidas e que o alinhamento dos substituintes seja evidente em casos de amostras que apresentem repetibilidade dos grupos. Essas imagens são os instrumentos fundamentais na criação da matriz de descritores moleculares, sendo então salvas em tamanhos iguais no formato BITMAP.

No primeiro tratamento de dados, o mapa de bit que computacionalmente gera a imagem é desdobrado em uma única linha para cada imagem, com as colunas representando as coordenadas originais dos pontos e preenchidas pelos valores correspondentes aos BIT's que o ocupavam, gerando, após a junção de todos os dados coletados, uma matriz única que dispõe em cada linha uma das amostras empregadas na análise. A matriz obtida em valores de BIT tem seus valores substituídos pelas propriedades periódicas dos átomos, sendo necessária a abordagem de imagens de estruturas com seus componentes em cores sólidas, visto que os efeitos de iluminação dos átomos resultam na identificação de um intervalo de tonalidades para um mesmo elemento. Esse efeito, por permitir valores de BIT diferentes a um mesmo átomo, torna inviável a tradução de todos os descritores de um elemento aos parâmetros químicos desejados.

A criação do vetor coluna com os valores de atividades biológicas é então realizado no bloco Y, pareando de modo a corresponder a propriedade às características estruturais das amostras dispostas no bloco X. A separação do conjunto empregado na validação externa é então realizada com a fração de dados utilizados para teste variando entre 20% a 30% da magnitude do conjunto total de dados avaliados (ALVES *et al.*, 2018), sendo o grupo de teste externo utilizado apenas na verificação da aplicabilidade do modelo gerado à predição de suas amostras, apresentando-se como conjunto independente do grupo de treinamento.

Para a separação do conjunto teste utilizado na validação externa do modelo, alguns métodos podem ser adotados para a seleção das amostras que o compõem, como por exemplo o Kennard-Stone e o randômico. O método Kennard-Stone (KENNARD; STONE, 1969) consiste na aplicação de um algoritmo capaz de separar amostras quanto a suas distâncias Euclidianas. Com isso, as amostras selecionadas apresentam maior distanciamento em termos de suas similaridades estruturais, resultando em um conjunto de treinamento mais abrangente

e uniforme (FERREIRA; TEIXEIRA; PETERNELLI, 2022), além de propor um grupo teste com amostras que apresentam características mais diversas. O método randômico, por sua vez, adota a amostragem probabilística do tipo aleatória simples (MARTÍNEZ-MESA *et al.*, 2016), buscando representantes de modo fortuito na composição dos grupos teste para a validação externa.

As amostras remanescentes do conjunto são utilizadas na construção da equação de regressão linear, gerada através de análise multivariada por métodos como o de mínimos quadrados parciais (PLS). O PLS é um método do tipo supervisionado que pode ser empregado no tratamento de conjuntos relativamente grandes, sendo indicado em situações em que seja necessário lidar com ruídos experimentais e colinearidades (FERREIRA *et al.*, 1999). Seu princípio visa a redução de dimensionalidade no tratamento de dados de determinado conjunto para a realização de uma análise com menor número de componentes, apresentando, como vantagem, o tratamento simultâneo das matrizes de descritores e de atividade biológica para a obtenção das variáveis latentes, responsáveis por descrever as frações que cada dimensão gerada detém de informações do conjunto de amostras empregado.

Na escolha da melhor quantidade de variáveis latentes utilizadas na construção do modelo, o gráfico de erro médio por variável latente é empregado, assumindo, preferencialmente, a menor quantidade para a descrição dentre as que apresentem menor valor de erro médio sem ganhos significativos de informação. A exigência de um modelo com baixo valor de variáveis latentes é justificado pelo superajuste causado na adoção de quantidades superiores às necessárias, podendo também incorporar dados obtidos por ruídos experimentais, que prejudicam a robustez do modelo.

Outro fator de grande importância na elaboração de um modelo QSAR é o tratamento estatístico, que afere o quão robusto e confiável ele é na descrição do conjunto de estruturas. Para que a equação possa ser adotada na descrição de determinado grupo de moléculas, a validação estatística deve ser realizada para garantir seu grau de confiabilidade em sua capacidade preditiva, sendo empregados métodos de validação e cálculo de parâmetros específicos na certificação de seu potencial, robustez, aptidão e estabilidade (CHATTERJEE; ROY, 2021). Estes métodos são incorporados em análises do tipo interna e externa.

Os métodos de validação interna são compreendidos como os que utilizam, para a manipulação dos dados e cálculos de parâmetros, as amostras presentes na construção do modelo matemático responsável pela predição. Inicialmente, alguns compostos do conjunto

de treinamento são selecionados para compor os dados para teste, com o método de validação interna cruzada aplicado para a obtenção do erro médio referente a cada variável latente. Esse método consiste na retirada de amostras do grupo de calibração para a criação de um modelo com as remanescentes e posterior predição das amostras isoladas, sendo realizados ciclos até que todas as amostras do conjunto de calibração componham ao menos uma vez o grupo teste da validação interna. As análises são feitas geralmente pelo *leave-one-out*, em que apenas uma amostra é retirada para compor o grupo teste e o *leave-many-out*, em que mais amostras são selecionadas (ALVES *et al.*, 2018). Outro método de validação interna empregado no método aug-MIA-QSAR é o de randomização do bloco y, em que as atividades biológicas até então pareadas às amostras são aleatorizadas e um novo modelo é gerado. Essa validação consegue estabelecer a correta dependência dos descritores moleculares com suas respectivas atividades, certificando a inexistência de efeitos de dados obtidos ao acaso. Para isso, espera-se o prejuízo do modelo gerado, evidenciando a existência de correlação entre os constituintes da matriz X e do vetor y através da obtenção de linearidades baixas e alto erro associado.

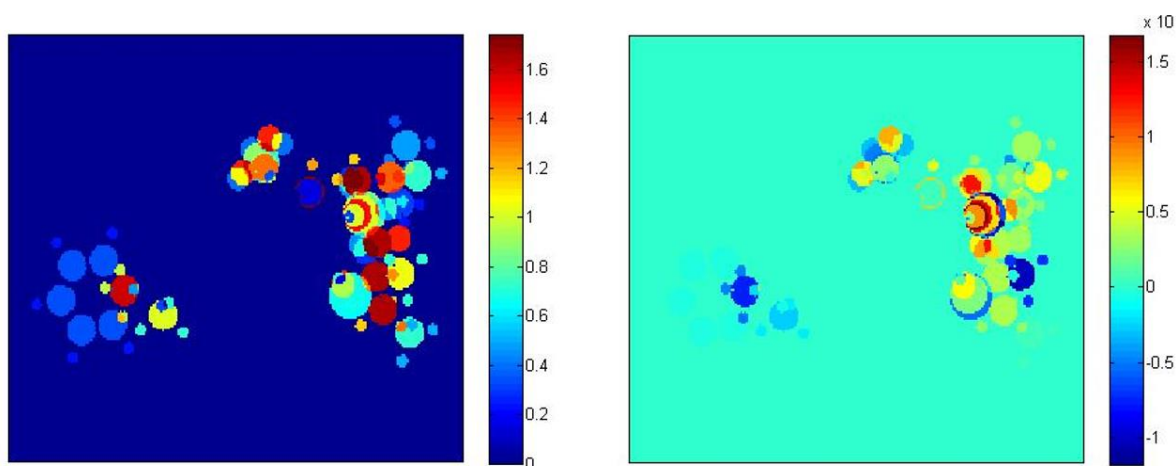
Os parâmetros estatísticos são então calculados e empregados na obtenção da robustez do modelo, sendo os mais clássicos para modelos QSAR o R^2 para validação externa e o Q^2 para validação interna (OJHA; ROY, 2018). O R^2 é o coeficiente de determinação, caracterizado por descrever a dispersão das variáveis preditas, creditando maior confiabilidade a modelos que assumam valor de R^2 mais próximo a 1 (TUKUR; SHALLANGWA; IBRAHIM, 2019). De modo análogo, quão mais próximo de 1 for o coeficiente de validação cruzada, o Q^2 , mais robusto será o modelo para descrever determinado sistema, sendo este parâmetro responsável por ceder informações sobre a relação entre o valor médio das atividades definidas e o valor predito pelo modelo confeccionado na descrição da amostras empregadas no teste (TUKUR; SHALLANGWA; IBRAHIM, 2019). Alguns parâmetros adicionais podem ainda ser empregados na análise, sendo esses propostos por Mitra *et al.* (2010) e conhecidos como ${}^cR_p^2$ e R_m^2 . O ${}^cR_p^2$ é o parâmetro responsável por fornecer informações sobre a diferença entre R^2 obtido na calibração e do R^2 obtido na randomização de Y, sendo esperado um valor acima de 0,5 para que o modelo seja considerado robusto. Já o R_m^2 avalia a correlação dos valores experimentais e preditos no conjunto teste indicando, quando superior a 0,5, que os valores experimentais absolutos e preditos são congruentes (ROY *et al.*, 2013). Em suma, para que o modelo gerado seja reconhecido como robusto e preditivo, os parâmetros estatísticos de validação devem atender aos seguintes requisitos limites: os parâmetros R_m^2 , R_p^2 , R_{pred}^2 e Q^2 devem apresentar valores

superiores a 0,5; para o R^2 , esperam-se valores superiores à 0,6; os valores de parâmetros randomizados devem ser menores que os correspondentes não randomizados e a diferença entre R^2 e Q^2 não é recomendada ser superior a 0,3 (VEERASAMY; RAJAK; JAIN; SIVADASAN *et al.*, 2011).

Além dos parâmetros estatísticos obtidos, o artigo de Unger e Hansch publicado na revista *Journal of Medicinal Chemistry* (UNGER; HANSCH, 1973) estabeleceu cinco regras a serem consideradas na escolha dos modelos matemáticos a serem propostos, padronizando os métodos QSAR até então utilizados, propondo condições de segurança aos resultados obtidos e orientando para a escolha dos que deveriam ser ditos como melhores modelos. As referidas regras são: 1) a escolha de variáveis independentes a serem testadas no modelo; 2) a validação estatística das variáveis selecionadas para a certificação da robustez do modelo; 3) o princípio da parcimônia, que justifica a escolha do modelo mais simples na presença de outros de mesmo efeito; 4) a presença de um número de compostos pelo menos cinco vezes maior que a quantidade de variáveis empregadas, visando a minimização da correlação por coincidência; 5) e que o modelo quantitativo seja consistente com o mecanismo de ação dos compostos testados (GAUDIO; ZANDONADE, 2001).

Como resultado da análise multivariada de imagem, os MIA Plots (Figura 16), imagens que descrevem os comportamentos dos descritores para a melhor interpretabilidade de um modelo, são obtidos para auxiliar na proposição de novos compostos. O gráfico de coeficientes de regressão PLS permite a verificação da influência positiva ou negativa de determinado grupo na resposta de sua atividade biológica, enquanto a imagem VIP (*variable importance in projection*) descreve, em termos de magnitude, a significância dos grupos para os valores de atividade verificados (BARIGYE *et al.*, 2016). Na interpretação dos MIA Plots para as imagens VIP, grupos funcionais com colorações mais próximas ao vermelho apresentam maior peso no cálculo das atividades biológicas, sendo portanto grupos altamente influentes no que se refere às capacidades das moléculas em comportar-se como bom ou mau inibidor. Para o gráfico de coeficientes de regressão PLS (gráfico **b**), grupos com colorações que tendem à cor vermelha apresentam contribuições positivas ao valor da atividade bioquímica observada, enquanto os que tendem a coloração azul trazem prejuízos e são, portanto, componentes indesejáveis em moléculas cujo objetivo é o aumento da atividade.

Figura 16 – Imagnes MIA Plots obtidas pelo método MIA-QSAR.



Legenda: Da esquerda para direita, Gráfico VIP e gráfico **b**.

Fonte: Do autor (2022).

Portanto, no que se refere ao desenvolvimento de novos herbicidas, o aug-MIA-QSAR traz grandes contribuições, permitindo a proposição de novas moléculas através de minuciosos estudos sobre as dependências de atividades biológicas com fatores estruturais e periódicos, apresentando recursos para uma melhor compreensão da atuação desses compostos que, geralmente, visam alterações na atividade enzimática através da sua inserção no sítio ativo.

3.5 Docking Molecular

A técnica de *docking* molecular originou-se em 1982 como grande recurso no planejamento de novos fármacos (BATISTA *et al.* 2022). Utilizando os princípios da mecânica molecular no cálculo de energias de um complexo enzimático, ela visa identificar interações de menor energia existentes entre um ligante e um sítio receptor através da modificação das posições e conformações do ligante (PAGADALA; SYED; TUSZYNSKI, 2017).

Por se tratar de uma técnica fundamentada na mecânica molecular, há a exigência de aplicação de uma função de *scoring*, que é classificada em funções de campo de força, empíricas e *knowledge-based* (WANG; LU; WANG, 2003). O campo de força é visto como a equação que rege todas as forças atuantes sobre os átomos que compõem um sistema,

avaliando de modo aproximado as energias referentes às posições de átomos quimicamente ligados ou não, através de parâmetros empíricos ou calculados previamente. Os átomos não ligados do sistema são tratados intermolecularmente pelos termos de van der Waals e eletrostático, com a inserção do potencial de Lennard-Jones e a função Coulombiana, respectivamente (SALMASO; MORO, 2018).

Atualmente, diversos campos de força estão disponíveis nos estudos *in silico* de sistemas biológicos a fim de atender as especificidades dos sistemas com melhor desempenho. A família dos campos de força AMBER desenvolvidos pelo grupo Petter Kollman (WANG; CIEPLAK; KOLLMAN, 2000), por exemplo, iniciou-se com o tratamento de proteínas e ácidos nucleicos (WANG *et al.*, 2004). A equação 3.5.1 é responsável por descrever o campo de força AMBER.

$$E_{total} = \sum_{\text{ligação}} K_r (r - r_{eq})^2 + \sum_{\text{ângulos}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{diedral}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (3.5.1)$$

Legenda: O primeiro termo refere-se a energia entre dois átomos ligados, o segundo ao ângulo de ligação entre três átomos ligados covalentemente, o terceiro ao ângulo diedro referente à torção de uma ligação e seu entorno e o último, as interações de Van der Waals e eletrostáticas.

Fonte: Modificado de WANG, CIEPLAK e KOLLMAN (2000).

A fim de contemplar condições específicas não descritas do sistema e representá-lo de forma coerente às observações experimentais, a parametrização da equação é realizada considerando o ambiente e as propriedades da estrutura empregada. Desta forma, alguns termos são adaptados, sendo atualmente conhecidas algumas séries com modificações do campo de força AMBER original que permitem atender particularidades de ligantes, sítios ativos enzimáticos e carboidratos.

O emprego da equação do campo de força, na técnica de *docking* molecular, é dado na obtenção da superfície de energia livre da interação ligante-receptor. Deste modo, o mínimo global de toda a superfície explorada corresponde a condição de menor energia, com maior probabilidade de ocorrência experimental. A robustez do resultado obtido é, portanto, condicionada ao software do cálculo, à modalidade do método e ao modelo tridimensional da macromolécula empregados.

Dentre os mecanismos de *docking* desenvolvidos, o MolDock demonstra-se promissor através da acurácia de 87%, superior a grande parte dos algoritmos similares (THOMSEN; CHRISTENSEN, 2006). Integrado à plataforma Molegro Virtual Docker que permite boa usabilidade nas operações⁷, o MolDock baseia-se em um algoritmo de busca híbrido que combina a otimização por evolução diferencial com o algoritmo de predição de cavidades, acrescido de um termo responsável por monitorar interações de ligações de hidrogênio no sistema e melhorias no termo eletrostático (THOMSEN; CHRISTENSEN, 2006).

No que se refere ao tipo do ancoramento, os tipos rígido, semiflexível e flexível são adotados. O tipo rígido, geralmente empregado para grandes sistemas tais como proteína-proteína, não apresenta alterações conformacionais no ligante e no receptor; o semiflexível, empregado no ancoramento de moléculas grandes e pequenas tais como proteínas ou ácidos nucleicos e ligantes, mantém a estrutura da macromolécula invariável e a do ligante flexível; o docking flexível, por fim, apresenta flexibilidade em ambos os constituintes ancorados, sendo mais empregado no tratamento de interações entre dois ligantes pequenos (SALMASO; MORO, 2018; TAO *et al.*, 2020).

Sistemas calculados em maiores graus de liberdade são mais hábeis em representar com fidelidade as condições do sistema biológico (BATISTA *et al.*, 2021). O tratamento flexível da macromolécula, porém, exige elevado custo computacional por considerar fatores adicionais (TAO *et al.*, 2020), incorporando aos cálculos de energia os comportamentos dos componentes da cadeia polipeptídica. Desta forma, na tentativa de conciliar bons resultados a baixos custos computacionais, a satisfação no uso do *docking* semiflexível torna-se evidente, trazendo boa performance no tratamento de sistemas enzimáticos através das liberdades rotacionais, translacionais e conformacionais aos ligantes (PAGADALA; SYED; TUSZYNSKI, 2017). Os algoritmos responsáveis por ceder ligantes nas diversas orientações espaciais são classificados como sistemáticos e estocásticos. Enquanto na categoria dos algoritmos sistemáticos encontram-se os métodos determinísticos, em que os intervalos de busca onde o modelo atua são previamente definidos e consequentemente limitados, os algoritmos estocásticos aplicam alterações aleatórias nas variáveis de estado até que sejam contemplados os critérios de término da análise, sendo consideráveis mais hábeis na obtenção de soluções representativas (BATISTA *et al.*, 2021).

⁷ MOLEXUS. **Molegro Virtual Docker**. Disponível em: <http://molexus.io/molegro-virtual-docker/>. Acesso em: 22 dez. 2022.

A escolha da macromolécula adequada também apresenta grande relevância na qualidade dos resultados obtidos. Quanto às estruturas tridimensionais obtidas experimentalmente, as oriundas das técnicas cristalografia de raio X, espectroscopia de RMN e microscopia eletrônica são predominantes no *Protein Data Bank* (PDB), compreendendo ao montante de 84%⁸ às desenvolvidas através da cristalografia de raio X. Este montante pode estar relacionado às desvantagens do emprego da técnica de RMN que apresenta elevado custo de implementação, complexibilidade de tratamento dos dados extraídos do sistema e baixa resolução na identificação dos resíduos, por operar em proteínas solubilizadas⁹.

Algumas métricas acerca destas estruturas, como por exemplo a resolução e valores R, servem como importante recurso na avaliação da qualidade dos modelos. A resolução diz respeito à qualidade dos dados coletados no cristal empregado na análise, aferindo em um intervalo de 1 a 3Å o nível de detalhes que o modelo apresenta. Neste parâmetro, valores absolutos menores são preferíveis pois indicam alta precisão nas informações obtidas do padrão de difração, o que reflete em acurácia no mapa de densidade de elétrons calculado e geram maior confiança na localização dos átomos na estrutura¹⁰.

Já o Valor R, referente ao pareamento do modelo atômico obtido a partir dos dados cristalográficos ao padrão de difração simulado baseado no modelo, apresenta valores típicos próximos a 0,20. Valores próximos de 0 evidenciam o ajuste perfeito de ambos os modelos pareados, sendo por vezes oriundos de refinamentos que trazem viés e irrepresentatividade aos dados empregados. Deste modo, o Valor Livre de R é utilizado para validar o refinamento aplicado, sendo este dado pela capacidade do treinamento com 90% dos dados do conjunto original já refinado predizer os dados ausentes. A diferença na magnitude entre o Valor R e Valor Livre de R indica a intensidade do ajuste causada pelo refinamento dos dados gerados¹¹.

Por fim, para a predição computacional das estruturas através da sequência primária de proteínas, os métodos de modelagem *ab initio* e por homologia podem ser empregados, sendo o *ab initio* menos utilizados devido à falta de dados precisos do processo de enovelamento de proteínas e alto custo computacional para a obtenção dos resultados (SANTOS FILHO;

⁸ RSCB PDB. **PDB Data Distribution by Experimental Method and Molecular Type**. Disponível em: <https://www.rcsb.org/stats/summary>. Acesso em: 22 dez. 2022.

⁹ Attila Ambrus. **Comparison of NMR and X-Ray Crystallography – As Methods of the Protein Structure Determination**. Disponível em: <https://www.cryst.bbk.ac.uk/pps97/assignments/projects/ambrus/html.htm#comp>. Acesso em 13/02/2023.

¹⁰ RSCB PDB. **Resolution**. Disponível em: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/resolution>. Acesso em: 22 dez. 2022.

¹¹ RSCB PDB. **R-value and R-free**. Disponível em: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/r-value-and-r-free>. Acesso em: 22 dez. 2022.

ALENCASTRO, 2003). Estes métodos são geralmente aplicados a casos em que a estrutura tridimensional da proteína de interesse é ainda desconhecida.

4 METODOLOGIA

4.1 Softwares Utilizados

Microsoft Excel versão 2111, 2016;
Gaussian Inc. GaussView 5.0.8, 2000-2008;
Gaussian Inc. Gaussian 09W versão 7.0, 1995-2009;
Microsoft Paint versão 11.2110.0.0, 2021;
Chemoface versão 1.65, 2021;
Notepad++ versão 8.1.9.3, 2021;
DiscoveryStudio;
Molegro Virtual Docker;
Site Molinspiration (<https://molinspiration.com>)

4.2 Métodos

Dois conjuntos formados por compostos análogos de 2*H*-benzo[*b*][1,4]oxazin-3(4*H*)-onas (LI *et al.*, 2018) e niacina-tricetonas (FU *et al.*, 2019a) e suas respectivas atividades biológicas em função de IC₅₀ foram inicialmente agrupados. No *Excel*, as atividades biológicas foram convertidas a pIC₅₀ (-log IC₅₀) e a região de identidade presente em ambos os conjuntos foi modelada no software *GaussView* e otimizada ao nível semiempírico AM1 através do software *Gaussian 09*.

Após a obtenção do ângulo de melhor visualização dos grupos substituintes, as imagens das estruturas em RGB com átomos em proporção ao raio de van der Waals foram capturadas na menor janela do *GaussView* (236 × 204 pixels) e transferidas ao *Microsoft Paint*, onde foram salvas no formato BITMAP.

O *Chemoface* foi então empregado para a geração da matriz de descritores que teve as colunas de variação nula descartadas. A matriz foi então submetida ao software *Notepad++* para a recodificação dos descritores, sendo geradas matrizes com valores referentes a eletronegatividade, raio de van der Waals e a razão raio de van der Waals por eletronegatividade dos átomos do conjunto.

Para a realização da regressão PLS das matrizes recodificadas com o valor de atividade em pIC_{50} referente às amostras, o *Chemoface* foi novamente utilizado, sendo realizada a amostragem randômica na escolha do grupo teste equivalente a 25% do conjunto total. Quinze ciclos foram realizados na escolha fortuita das amostras, a média dos parâmetros estatísticos foi calculada e os gráficos **b** e **VIP** do modelo mais robusto foi analisado para a proposição das novas moléculas. A robustez do modelo foi avaliada pelo parâmetro %scores, em que os valores médios dos resultados estatísticos esperados acima de 0,5 são calculados e é visto como melhor o que apresenta valor mais próximo de 1,0.

Os modelos mais robustos de cada propriedade periódica foram empregados na predição das atividades em pIC_{50} das moléculas propostas e a média dos três valores obtidos para cada amostra foi realizada.

O conjunto final composto por todas as moléculas adquiridas foi submetido à técnica de *docking* molecular. Para isso, cada estrutura foi projetada no software *Discovery Studio* e submetida ao tratamento de minimização de energia, correção de duplas ligações e cálculo de cargas parciais, salvas posteriormente no formato mol2.

A estrutura cristalina da HPPD da erva *Arabidopsis Thaliana* (AtHPPD, EC1.13.11.27) complexada com o inibidor 1,5-(dimetil-6-(2-oxidanil-6-oxidanilideno-ciclohexen-1-il) carbonil-3-(3-trimetilsiliprop-2-inil)quinazilina-2,4,diona (NJX) foi adquirida no banco de dados (PDB ID: 6M6D; resolução:1,80 Å) e os tratamentos de inserção de hidrogênios e retirada de moléculas de água (exceto H₂O664) foram feitos. As cargas eletrostáticas foram definidas pelo campo de força AMBER e a estrutura foi refinada usando algoritmos de gradientes conjugados até que o gradiente de energia da raiz quadrada média final fosse inferior a 0,10 kcal mol⁻¹.

Para a remoção de contatos próximos que possam ter sido negligenciados na adição de hidrogênios, realizou-se o tratamento de restrições e relaxamentos em toda a estrutura até que toda molécula fosse relaxada e os ligantes foram otimizados com o campo de força AMBER.

O software *Molegro Virtual Docker* foi então utilizado na realização dos cálculos de *docking* molecular. A região escolhida consistiu em uma grid cúbica, com resolução de 0,3 Å e no raio de 10 Å a partir do inibidor. O *MolDock SE* foi o algoritmo de busca usado na obtenção das posições do ligante, com número máximo de poses igual a 100 como requisito de término e as regiões de ligação em potencial foram identificadas por algoritmo de detecção de cavidade, sendo essas monitoradas no processo de ancoramento.

Por fim, as estruturas foram projetadas no site *Molinspiration*, os valores correspondentes ao $\log P$ de cada estrutura foram analisados e os mesmos procedimentos foram aplicados às moléculas análogas de 2-ariloxiacetilciclohexan-1,3-dionas (WANG *et al.*, 2016) e quinolina-tricetonas (WANG *et al.*, 2015), cuja atividade biológica foi dada em função de K_i .

REFERÊNCIAS

ALVES, V. M. *et al.* QUIMIOINFORMÁTICA: UMA INTRODUÇÃO. **Química Nova**, [São Paulo], v. 41, n. 2, p. 202-212, 2018. DOI: 10.21577/0100-4042.20170145.

BARIGYE, S. J. *et al.* MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. **RSC Advances**, [London], v. 6, n. 55, p. 49604-49612, 2016. DOI: 10.1039/C6RA09593C.

BATISTA, V. *et al.* CONSTRUÇÃO, OTIMIZAÇÃO E ANCORAGEM MOLECULAR DE SUBSTÂNCIAS BIOATIVAS EM BIOMACROMOLÉCULAS: UM TUTORIAL PRÁTICO. **Química Nova**, [São Paulo], v. 45, n. 2, p. 223-234, 2021. DOI: 10.21577/0100-4042.20170821.

BEAUDEGNIES, R. *et al.* Herbicidal 4-hydroxyphenylpyruvate dioxygenase inhibitors—A review of the triketone chemistry story from a Syngenta perspective. **Bioorganic & Medicinal Chemistry**, [Oxford], v. 17, n. 12, p. 4134-4152, 2009. DOI: 10.1016/j.bmc.2009.03.015.

BENFENATI, E. *et al.* Predicting logP of pesticides using different software. **Chemosphere**, [Oxford], v. 53, n. 9, p. 1155-1164, 2003. DOI: 10.1016/S0045-6535(03)00609-X.

BRAUSEMANN, A. *et al.* Structure of Phytoene Desaturase Provides Insights into Herbicide Binding and Reaction Mechanisms Involved in Carotene Desaturation. **Structure**, [Cambridge], v. 25, n. 8, p. 1222-1232, 2017. DOI: 10.1016/j.str.2017.06.002.

BURLINGHAM, B. T.; WIDLANSKI, T. S. An Intuitive Look at the Relationship of K_i and IC_{50} : A More General Use for the Dixon Plot. **Journal of Chemical Education**, [Washington DC] v. 80, n. 2, p. 214-218, 2003. DOI: 10.1021/ed080p214.

CASIDA, J. E. Pesticide interactions: mechanisms, benefits, and risks. **Journal of agricultural and food chemistry**, [Washington DC], v. 65, n. 23, p. 4553-4561, 2017. DOI: 10.1021/acs.jafc.7b01813.

CERQUEIRA, F. M.; MEDEIROS, M. H. G. D.; AUGUSTO, O. Antioxidantes dietéticos: controvérsias e perspectivas. **Química Nova**, [São Paulo], v. 30, n. 2, p. 441-449, 2007. DOI: 10.1590/S0100-40422007000200036.

CHAHAL, P. S.; JUGULAM, M.; JHALA, A. J. Basis of Atrazine and Mesotrione Synergism for Controlling Atrazine- and HPPD Inhibitor-Resistant Palmer Amaranth. **Agronomy Journal**, [Madison], v. 111, n. 6, p. 3265-3273, 2019. DOI: 10.2134/agronj2019.01.0037.

CHATTERJEE, M.; ROY, K. Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors. **Journal of Hazardous Materials**, [Netherlands], v. 408, p. 1-13, 2021. DOI: 10.1016/j.jhazmat.2020.124936.

CONSONNI, V.; TODESCHINI, R. Molecular Descriptors. *In: PUZYN, T. et al. (Ed.). Recent Advances in QSAR Studies: Methods and Applications.* Netherlands: Springer, 2010.

CORMANICH, R. A.; NUNES, C. A.; FREITAS, M. P. Desenhos de estruturas químicas correlacionam-se com propriedades biológicas: MIA-QSAR. **Química Nova**, [São Paulo], v. 35, n. 6, p. 1157-1163, 2012. DOI: 10.1590/S0100-40422012000600017.

CORNES, D. Callisto: a very successful maize herbicide inspired by allelochemistry. *In: WORLD CONGRESS ON ALLELOPATHY*, 4, 2005, Wagga Wagga. **Anais [...]**. Wagga Wagga: Centre of Rural Social Research, 2005, p. 569-572.

COSTA, N. V. *et al.* Métodos de controle de plantas daninhas em sistemas orgânicos: breve revisão. **Revista Brasileira de Herbicidas**, [Londrina], v. 17, n. 1, p. 25-44, 2018. DOI: 10.7824/rbh.v17i1.522.

CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, [Washington DC], v. 110, n. 18, p. 5959-5967, 1988. DOI: 10.1021/ja00226a005.

DAYAN, F. E. *et al.* beta-Triketone Inhibitors of Plant p-Hydroxyphenylpyruvate Dioxygenase: Modeling and Comparative Molecular Field Analysis of Their Interactions. **Journal of Agricultural and Food Chemistry**, [Washington DC], v. 57, n. 12, p. 5194-5200, 2009. DOI: 10.1021/jf9005593.

DELLAPENNA, D.; POGSON, B. Vitamin Synthesis in Plants: Tocopherols and Carotenoids. **Annual review of plant biology**, [United States], v. 57, p. 711-738, 2006. DOI: 10.1146/annurev.arplant.56.032604.144301.

DEWAR, M. J. S. *et al.* Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. **Journal of the American Chemical Society**, [United States], v. 107, n. 13, p. 3902-3909, 1985. DOI: 10.1021/ja00299a024.

DIXON, D. P.; EDWARDS, R. Enzymes of tyrosine catabolism in *Arabidopsis thaliana*. **Plant science**, [Ireland], v. 171, n. 3, p. 360-366, 2006. DOI: 10.1016/j.plantsci.2006.04.008.

DUARTE, M.H. **INOVAÇÃO DO MÉTODO MIA-QSAR: APLICAÇÃO PARA DOENÇAS TROPICAIS NEGLIGENCIADAS.** Lavras: Ed. Lavras, 2016. *online*.

DUMAS, E. *et al.* Fate and ecotoxicological impact of new generation herbicides from the triketone family: An overview to assess the environmental risks. **Journal of Hazardous Materials**, [Netherlands], v. 325, p. 136-156, 2017. DOI: 10.1016/j.jhazmat.2016.11.059.

DYSON, J. S. *et al.* Adsorption and Degradation of the Weak Acid Mesotrione in Soil and Environmental Fate Implications. **Journal of Environmental Quality**, [United States], v. 31, n. 2, p. 613-618, 2002. DOI: 10.2134/jeq2002.6130.

EROS, D. *et al.* Reliability of logP predictions based on calculated molecular descriptors: a critical review. **Current Medicinal Chemistry**, [United Arab Emirates], v. 9, n. 20, p. 1819-1829, 2002. DOI: 10.2174/0929867023369042.

FERREIRA, M. M. C. *et al.* Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, [São Paulo], v. 22, n. 5, p. 724-731, 1999. DOI: 10.1590/S0100-40421999000500016.

FERREIRA, R. D. A.; TEIXEIRA, G.; PETERNELLI, L. A. Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data. **Ciência Rural**, [Brazil], v. 52, n. 5, 2022. DOI: 10.1590/0103-8478cr20201072.

FREITAS, M. P.; BROWN, S. D.; MARTINS, J. A. MIA-QSAR: a simple 2D image-based approach for quantitative structure–activity relationship analysis. **Journal of Molecular Structure**, [Netherlands], v. 738, n. 1, p. 149-154, 2005. DOI: 10.1016/j.molstruc.2004.11.065.

FREITAS, M. P.; RAMALHO, T. C. Employing conformational analysis in the molecular modeling of agrochemicals: Insights on QSAR parameters of 2,4-D. **Ciencia e Agrotecnologia**, [Brazil], v. 37, n. 6, p. 485-494, 2013. DOI: 10.1590/S1413-70542013000600001.

FREITAS, M. R.; BARIGYE, S. J.; FREITAS, M. P. Coloured chemical image-based models for the prediction of soil sorption of herbicides. **RSC Advances**, [United Kingdom], v. 5, n. 10, p. 7547-7553, 2015. DOI: 10.1039/C4RA12070A.

FU, Q. *et al.* Synthesis and herbicidal activity of novel pyrazole aromatic ketone analogs as HPPD inhibitor. **Pest Management Science**, [United Kingdom], v. 76, n. 3, p. 868-879, 2020. DOI: 10.1002/ps.5591.

FU, Y. *et al.* 3D pharmacophore-based virtual screening and docking approaches toward the discovery of novel HPPD inhibitors. **Molecules**, [Switzerland], v. 22, n. 6, p. 959, 2017a. DOI: 10.3390/molecules22060959.

FU, Y. *et al.* Design, synthesis, herbicidal activity and CoMFA of aryl-formyl piperidinone HPPD inhibitors. **Pesticide Biochemistry and Physiology**, [United States], v. 174, p. 104811, 2021. DOI: 10.1016/j.pestbp.2021.104811.

FU, Y. *et al.* Design, synthesis, SAR and molecular docking of novel green niacin-triketone HPPD inhibitor. **Industrial Crops and Products**, [Netherlands], v. 137, p. 566-575, 2019a. DOI: 10.1016/j.indcrop.2019.05.070.

FU, Y. *et al.* Discovery of N-Aroyl Diketone/Triketone Derivatives as Novel 4-Hydroxyphenylpyruvate Dioxygenase Inhibiting-Based Herbicides. **Journal of Agricultural and Food Chemistry**, [Washington DC], v. 67, n. 43, p. 11839-11847, 2019b. DOI: 10.1021/acs.jafc.9b01412.

FU, Y. *et al.* Identification of novel inhibitors of p-hydroxyphenylpyruvate dioxygenase using receptor-based virtual screening. **Journal of the Taiwan Institute of Chemical Engineers**, [Taiwan], v. 103, p. 33-43, 2019c. DOI: 10.1016/j.jtice.2019.08.005.

FU, Y. *et al.* New Research for Quinazoline-2,4-diones as HPPD Inhibitors Based on 2D-MLR and 3D-QSAR Models. **Combinatorial Chemistry & High Throughput Screening**, [United Arab Emirates], v. 20, n. 9, p. 748-759, 2017b. DOI: 10.2174/1386207320666170622073738.

FU, Y. *et al.* Quantitative structure activity relationship studies and molecular dynamics simulations of 2-(aryloxyacetyl) cyclohexane-1, 3-diones derivatives as 4-hydroxyphenylpyruvate dioxygenase inhibitors. **Frontiers in chemistry**, [Switzerland], v. 7, p. 556, 2019d. DOI: 10.3389/fchem.2019.00556.

GALLI, F. Vitamin E-derived copolymers continue the challenge to hemodialysis biomaterials. **World Journal of Nephrology**, [*s. l.*], v. 1, n. 4, p. 100-105, 2012. DOI: 10.5527/wjn.v1.i4.100.

GAUDIO, A. C.; ZANDONADE, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Química Nova**, [São Paulo], v. 24, n. 5, p. 658-671, 2001. DOI: 10.1590/S0100-40422001000500013.

GUARATINI, T.; MEDEIROS, M. H. G.; COLEPICOLO, P. Antioxidantes na manutenção do equilíbrio redox cutâneo: uso e avaliação de sua eficácia. **Química Nova**, [São Paulo], v. 30, n. 1, p. 206-213, 2007. DOI: 10.1590/S0100-40422007000100033.

HANSCH, C.; FUJITA, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. **Journal of the American Chemical Society**, [United States], v. 86, n. 8, p. 1616-1626, 1964. DOI: 10.1021/ja01062a035.

HAWKES, T. R. *et al.* Characterization of 4-hydroxyphenylpyruvate dioxygenases, inhibition by herbicides and engineering for herbicide tolerance in crops. **Pesticide Biochemistry and Physiology**, [United States], v. 156, p. 9-28, 2019. DOI: 10.1016/j.pestbp.2019.01.006.

HE, B. *et al.* Discovery of Novel Pyrazole-Quinazoline-2,4-dione Hybrids as 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. **Journal of Agricultural and Food Chemistry**, [Washington DC], v. 68, n. 18, p. 5059-5067, 2020. DOI: 10.1021/acs.jafc.0c00051.

HUANG, J. J. *et al.* HerbiPAD: a free web platform to comprehensively analyze constitutive property and herbicide-likeness to estimate chemical bioavailability. **Pest Management Science**, [United Kingdom], v. 77, n. 3, p. 1273-1281, 2021. DOI: 10.1002/ps.6140.

JACOBS, K. E. *et al.* Association between metabolic resistances to atrazine and mesotrione in a multiple-resistant waterhemp (*Amaranthus tuberculatus*) population. **Weed Science**, [United States], v. 68, n. 4, p. 358-366, 2020. DOI: 10.1017/wsc.2020.31.

KARAM, D. *et al.* Situação atual da resistência de plantas daninhas a herbicidas nos sistemas agrícolas. In: CONGRESSO NACIONAL DE MILHO E SORGO, 32, 2018, Lavras. **Anais [...]**. Sete Lagoas: Associação Brasileira de Milho e Sorgo, 2018. p. 901-932.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, [United Kingdom], v. 11, n. 1, p. 137-148, 1969. DOI: 10.2307/1266770.

KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. **Journal of Medicinal Chemistry**, [United States], v. 37, n. 24, p. 4130-4146, 1994. DOI: 10.1021/jm00050a010.

KOHLHASE, D. R.; EDWARDS, J. W.; OWEN, M. D. K. Inheritance of 4-hydroxyphenylpyruvate dioxygenase inhibitor herbicide resistance in an *Amaranthus tuberculatus* population from Iowa, USA. **Plant science**, [Limerick], v. 274, p. 360-368, 2018. DOI: 10.1016/j.plantsci.2018.06.004.

KUPPER, A. *et al.* Tembotrione detoxification in 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitor-resistant Palmer amaranth (*Amaranthus palmeri* S. Wats.). **Pest Management Science**, [United Kingdom], v. 74, n. 10, p. 2325-2334, 2018. DOI: 10.1002/ps.4786.

LI, H. B. *et al.* Novel HPPD inhibitors: triketone 2H-benzo b 1,4 oxazin-3(4H)-one analogs. **Pest Management Science**, [United Kingdom], v. 74, n. 3, p. 579-589, 2018. DOI: 10.1002/ps.4739.

MARTÍNEZ-MESA, J. *et al.* Sampling: how to select participants in my research study? **Anais Brasileiros de Dermatologia**, [Brazil], v. 91, n. 3, p. 326-330, 2016. DOI: 10.1590/abd1806-4841.20165254.

MESQUITA, S.; TEIXEIRA, C.; SERVULO, E. Carotenoids: Properties, Applications and Market. **Revista Virtual de Química**, [Brazil], v. 9, n. 2, p. 672-688, 2017. DOI: 10.21577/1984-6835.20170040.

MOLCHANOV, S.; GRYFF-KELLER, A. Inhibition of 4-hydroxyphenylpyruvate dioxygenase by 2-[2-nitro-4-(trifluoromethyl)benzoyl]-1,3-cyclohexanedione. **Acta Biochimica Polonica**, [Poland], v. 56, n. 3, p. 447-454, 2009. PMID: 19621095.

MORAN, G. R. 4-Hydroxyphenylpyruvate dioxygenase. **Archives of Biochemistry and Biophysics**, [United States], v. 433, n. 1, p. 117-128, 2005. DOI: 10.1016/j.abb.2004.08.015.

NDIKURYAYO, F. *et al.* 4-Hydroxyphenylpyruvate dioxygenase inhibitors: From chemical biology to agrochemicals. **Journal of Agricultural and Food Chemistry**, [Washington DC], V. 65, n. 39, p. 8523-8537, 2017. DOI: 10.1021/acs.jafc.7b03851.

NORRIS, S. R.; BARRETTE, T. R.; DELLAPENNA, D. Genetic dissection of carotenoid synthesis in *Arabidopsis* defines plastoquinone as an essential component of phytoene desaturation. **The Plant Cell**, [United States], v. 7, n. 12, p. 2139-2149, 1995. DOI: 10.1105/tpc.7.12.2139.

O'BRIEN, S. R.; DAVIS, A. S.; RIECHERS, D. E. Quantifying Resistance to Isoxaflutole and Mesotrione and Investigating Their Interactions with Metribuzin POST in Waterhemp (*Amaranthus tuberculatus*). **Weed Science**, [United States], v. 66, n. 5, p. 586-594, 2018. DOI: 10.1017/wsc.2018.36.

OJHA, P. K.; ROY, K. Development of a robust and validated 2D-QSPR model for sweetness potency of diverse functional organic molecules. **Food and Chemical Toxicology**, [United Kingdom], v. 112, p. 551-562, 2018. DOI: 10.1016/j.fct.2017.03.043.

OLIVEIRA JR., R. S. D.; CONSTANTIN, J.; INOUE, M. H. **Biologia e Manejo de Plantas Daninhas**. 1 ed. Curitiba: Omnipax, 2011.

PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. **Biophysical Reviews**, [Germany], v. 9, n. 2, p. 91-102, 2017. DOI: 10.1007/s12551-016-0247-1.

PENG, Y. J. *et al.* Quintrione: A new selective herbicide for weed control in rice (*Oryza sativa* L.). **Crop Protection**, [United Kingdom], v. 141, 2021. DOI: 10.1016/j.cropro.2020.105501.

PYKA, A.; BABUŚKA, M.; ZACHARIASZ, M. A comparison of theoretical methods of calculation of partition coefficients for selected drugs. **Acta Poloniae Pharmaceutica**, [Poland], v. 63, n. 3, p. 159-167, 2006. PMID: 20085219.

QU, R. Y. *et al.* Structure-Guided Discovery of Silicon-Containing Subnanomolar Inhibitor of Hydroxyphenylpyruvate Dioxygenase as a Potential Herbicide. **Journal of Agricultural and Food Chemistry**, [Washington DC], v. 69, n. 1, p. 459-473, 2021. DOI: 10.1021/acs.jafc.0c03844 .

ROMAN, E.S. *et al.* **Como funcionam os herbicidas: da biologia à aplicação**. Passo Fundo: Berthier, 2007.

ROY, K.; AMBURE, P.; AHER, R. B. How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models? **Chemometrics and Intelligent Laboratory Systems**, [Netherlands], v. 162, p. 44-54, 2017. DOI: 10.1016/j.chemolab.2017.01.010.

ROY, K. *et al.* Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: Emphasis on scaling of response data. **Journal of Computational Chemistry**, [United States], v. 34, n. 12, p. 1071-1082, 2013. DOI: 10.1002/jcc.23231.

SALMASO, V.; MORO, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. **Frontiers in Pharmacology**, [Switzerland], v. 9, p. 923, 2018. DOI: 10.3389/fphar.2018.00923.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. D. Modelagem de proteínas por homologia. **Química Nova**, [São Paulo], v. 26, n. 2, p. 253-259, 2003. DOI: 10.1590/S0100-40422003000200019.

SANTUCCI, A. *et al.* 4-Hydroxyphenylpyruvate Dioxygenase and Its Inhibition in Plants and Animals: Small Molecules as Herbicides and Agents for the Treatment of Human Inherited Diseases. **Journal of Medicinal Chemistry**, [United States], v. 60, n. 10, p. 4101-4125, 2017. DOI: 10.1021/acs.jmedchem.6b01395.

SETH, A.; OJHA, P. K.; ROY, K. QSAR modeling with ETA indices for cytotoxicity and enzymatic activity of diverse chemicals. **Journal of Hazardous Materials**, [Netherlands], v. 394, 2020. DOI: 10.1016/j.jhazmat.2020.122498.

SHAH, D. D. *et al.* Evidence for the mechanism of hydroxylation by 4-hydroxyphenylpyruvate dioxygenase and hydroxymandelate synthase from intermediate partitioning in active site variants. **Biochemistry**, [United States], v. 50, n. 35, p. 7694-7704, 2011. DOI: 10.1021/bi2009344.

SILVA, P. V. D.; MONQUERO, P. A.; MUNHOZ, W. S. CONTROLE EM PÓS-EMERGÊNCIA DE PLANTAS DANINHAS POR HERBICIDAS UTILIZADOS NA CULTURA DA CANA-DE-AÇÚCAR. **Revista Caatinga**, [Brasil], v. 28, n. 4, p. 21-32, 2015. DOI: 10.1590/1983-21252015v28n403rc.

SUN, Y. *et al.* Dissipation and Dynamics of Mesotrione in Maize and Soil Under Field Ecosystem. **Bulletin of environmental contamination and toxicology**, [United States], v. 90, n. 2 2012. DOI: 10.1007/s00128-012-0907-1.

TAO, X. *et al.* Recent developments in molecular docking technology applied in food science: a review. **International journal of food science & technology**, [United Kingdom], v. 55, n. 1, p. 33-45, 2020. DOI: 10.1111/ijfs.14325.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. **Journal of Medicinal Chemistry**, [United States], v. 49, n. 11, p. 3315-3321, 2006. DOI: 10.1021/jm051197e.

TUCCI, M. *et al.* A Storable Mediatorless Electrochemical Biosensor for Herbicide Detection. **Microorganisms**, [Switzerland], v. 7, n. 12, p. 630, 2019. DOI: 10.3390/microorganisms7120630.

TUKUR, S.; SHALLANGWA, G. A.; IBRAHIM, A. Theoretical QSAR modelling and molecular docking studies of some 4-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme inhibitors potentially used as herbicides. **Heliyon**, [Netherlands], v. 5, n. 11, p. 2859, 2019. DOI: 10.1016/j.heliyon.2019.e02859.

UNGER, S. H.; HANSCH, C. On model building in structure-activity relationships. A reexamination of adrenergic blocking activity of beta-halo-beta-arylalkylamines. **Journal of Medicinal Chemistry**, [United States], v. 16, n. 7, p. 745-749, 1973. DOI: 10.1021/jm00265a001.

VEERASAMY, R. *et al.* Validation of QSAR Models-Strategies and Importance. **International Journal of Drug Design and Discovery**, [*s. l.*], v. 2, n. 3, p. 511-519.

WANG, D. W. *et al.* Synthesis and Herbicidal Activity of Triketone-Quinoline Hybrids as Novel 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. **Journal of Agriculture and Food Chemistry**, [Washington DC], v. 63, n. 23, p. 5587-5596, 2015. DOI: 10.1021/acs.jafc.5b01530.

WANG, D. W. *et al.* An Efficient One-Pot Synthesis of 2-(Aryloxyacetyl)cyclohexane-1,3-diones as Herbicidal 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. **Journal of**

Agricultural and Food Chemistry, [Washington DC], v. 64, n. 47, p. 8986-8993, 2016. DOI: 10.1021/acs.jafc.6b04110.

WANG, J.; CIEPLAK, P.; KOLLMAN, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? **Journal of Computational Chemistry**, [United States], v. 21, n. 12, p. 1049-1074, 2000. DOI: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F.

WANG, J. *et al.* Development and testing of a general amber force field. **Journal of Computational Chemistry**, [United States], v. 25, n. 9, p. 1157-1174, 2004. DOI: 10.1002/jcc.20035.

WANG, R.; LU, Y.; WANG, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. **Journal of Medicinal Chemistry**, [United States], v. 46, n. 12, p. 2287-2303, 2003. DOI: 10.1021/jm0203783.

WANG, W. *et al.* Ionic Liquid Forms of Mesotrione with Enhanced Stability and Reduced Leaching Risk. **ACS Sustainable Chemistry & Engineering**, [United States], v. 7, n. 19, p. 16620-16628, 2019. DOI: 10.1021/acssuschemeng.9b03948.

WANG, Z. W. *et al.* Fragments recombination, design, synthesis, safener activity and CoMFA model of novel substituted dichloroacetylphenyl sulfonamide derivatives. **Pest Management Science**, [United Kingdom], v. 77, n. 4, p. 1724-1738, 2021. DOI: 10.1002/ps.6193.

XU, J. J. *et al.* General and specialized tyrosine metabolism pathways in plants. **aBIOTECH**, [s. l.], v. 1, n. 2, p. 97-105, 2020. DOI: 10.1007/s42994-019-00006-w.

ZIMDAHL, R. L. Chapter 13 - Introduction to Chemical Weed Control. *In*: ZIMDAHL, R. L. **Fundamentals of Weed Science**. 5 ed. Cambridge: Academic Press, 2018a. p. 391-416.

ZIMDAHL, R. L. Chapter 16 - Properties and Uses of Herbicides. *In*: ZIMDAHL, R. L. **Fundamentals of Weed Science**. 5 ed. Cambridge: Academic Press, 2018b. p. 463-499.

ZIMDAHL, R. L. Chapter 19 - The Problem and Study of Herbicide Resistance. *In*: ZIMDAHL, R. L. **Fundamentals of Weed Science**. 5 ed. Cambridge: Academic Press, 2018c. p. 535-555.

SEGUNDA PARTE

Artigo 1 – Normas do periódico Journal of Science of Food and Agriculture (versão preliminar)

MIA-QSAR and docking of recent HPPD inhibitors for predicting enzyme activity

Running Title: Triketone inhibitors of HPPD

Luiz R. Capucho,^a Ingrid V. Pereira,^a Adriana C. de Faria,^a Joyce K. Daré,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

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

* Corresponding author: matheus@ufla.br

Abstract

BACKGROUND: Mesotrione is a triketone herbicide widely used as an inhibitor of the hydroxyphenylpyruvate deoxygenase (HPPD) enzyme. However, new agrochemicals should be continuously developed to tackle the problem of herbicide resistance. Two sets of triketone analogs of mesotrione have been recently synthesized and demonstrated successful phytotoxicity against weeds. Herein, these compounds were joined to form a single data set and the HPPD inhibition of this enlarged library of triketones was modeled using multivariate image analysis applied to quantitative structure-activity relationships (MIA-QSAR). Docking studies were further carried out to validate the MIA-QSAR findings and aid the interpretation of ligand-enzyme interactions responsible for the bioactivities (pIC₅₀). **RESULTS:** The MIA-QSAR models based on van der Waals radii (r_{vdW}), electronegativity (ϵ), and r_{vdW}/ϵ ratio as molecular descriptors were similarly and satisfactorily predictive ($r^2 \geq 0.80$, $q^2 \geq 0.68$, and $r^2_{pred} \geq 0.68$). Subsequently, the PLS regression parameters were applied to predict the pIC₅₀ values of newly proposed derivatives, yielding a few promising agrochemical candidates. In addition, the calculated log P for most of these derivatives was found to be higher than that of mesotrione and the library compounds, indicating that they should be less prone to leach out and contaminate groundwater. **CONCLUSION:** MIA descriptors corroborated by docking studies were capable to reliably model the herbicidal activities of 68 triketones. Due to the substituent effects at the triketone framework, particularly of a nitro group in R³, promising analogs could be designed. The **P9** proposal demonstrated higher calculated activity and log P than the commercial mesotrione.

Keywords: QSAR; docking; triketones; HPPD enzyme

1. Introduction

Leptospermone (Figure 1) is a triketone obtained from myrtle species that was first isolated by Penfold in 1926.¹ It is a milestone in the inhibition of the hydroxyphenylpyruvate deoxygenase (HPPD) enzyme and, therefore, it has been employed as an herbicide since the 1970s.² Its structurally similar analog mesotrione (Figure 1) was then developed and is currently a major commercial agrochemical, widely used as a benchmark in development studies, especially in maize.

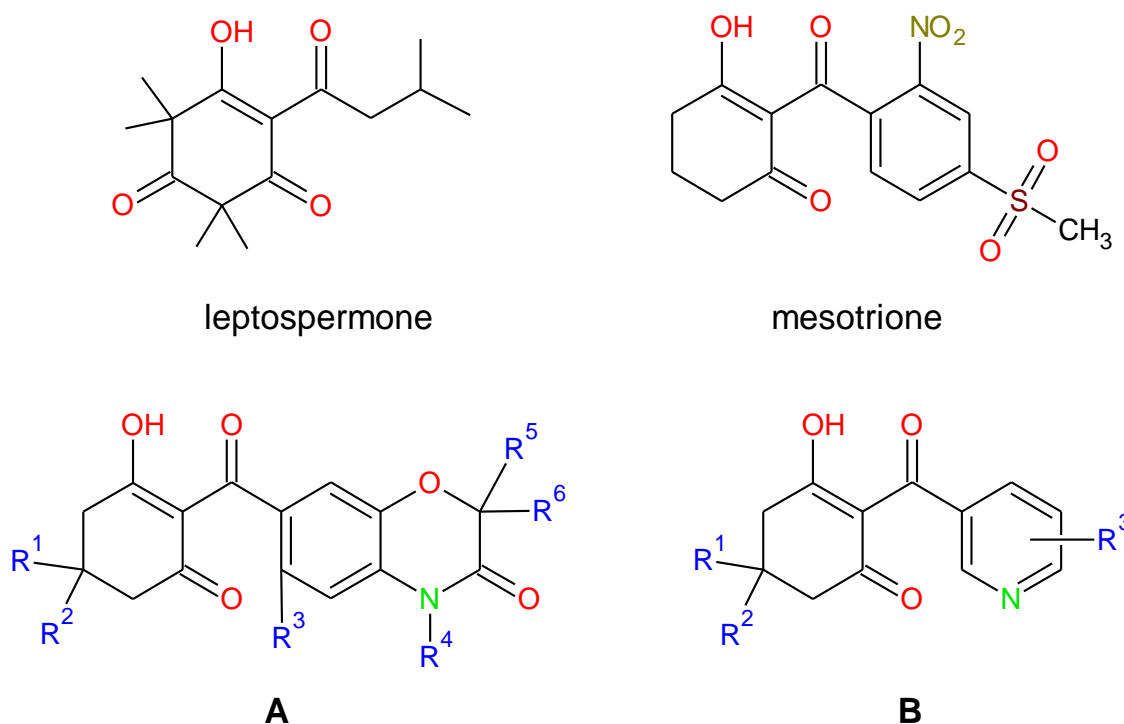


Figure 1. Chemical structures of leptospermone, mesotrione, and the analogs **A** and **B** studied herein.

Triketones like mesotrione act by blocking the HPPD enzyme, which is responsible to metabolize the *L*-tyrosine amino acid. This amino acid participates in the release of essential cofactors, such as plastoquinone and α -tocopherol, from homogentisate.³ These cofactors allow for the correct functioning of photosynthesis, the synthesis of carotenoids, and, most importantly, the inhibition of photo-oxidative processes of reactive species to the thylakoid membrane.⁴

Recently, two series of triketone analogs have been synthesized and tested as HPPD inhibitors (Figure 1), demonstrating potential application as herbicides in comparison to mesotrione.^{2,5} Since these compounds correspond to congeneric molecules and the biological assays were performed in the same manner, and also due to the almost identical IC₅₀ values

obtained for the reference compound (mesotrione) in both studies, they were combined to a single data set, thus providing an enlarged library of compounds for further quantitative structure-activity relationship (QSAR) studies. Multivariate image analysis (MIA) descriptors⁶ were applied to correlate the molecular structures with the corresponding pIC_{50} values of 68 triketones, while MIA contour maps of PLS regression coefficients and variable importance in projection scores⁷ were used to evaluate the substituent effects on the biological activity. Docking studies were used to understand the interaction mechanism between the proposed compounds and the HPPD enzyme. Accordingly, agrochemical candidates were designed to guide organic synthesis and tackle the problem of resistance to the available herbicides of this class of inhibitors, which is an ongoing problem in agriculture.^{8,9}

2. Materials and methods

The MIA-QSAR method was applied to study 68 molecules available elsewhere,^{2,5} which consist of 42 triketone 2*H*-benzo[*b*][1,4]oxazin-3(4*H*)-ones (**A**), 25 niacin-triketones (**B**), and mesotrione. The HPPD inhibition is described in terms of pIC_{50} (IC_{50} in mol L^{-1}). The molecules were sketched using the GaussView program,¹⁰ keeping the congruent moieties aligned (Figure 2). Atoms were represented as circles with sizes proportional to the respective van der Waals radii. The images of 204×236 pixels each were unfolded to row vectors and then grouped to yield the $67 \times 48,144$ matrix (**X**). The pixels are digital objects numbered according to the RGB scale, that is, from zero (black) to 765 (white). To add chemical meaning to these images, the pixels for each atom were renumbered to make a parallelism with the corresponding van der Waals radius (r_{vdW}), electronegativity (ϵ), and r_{vdW}/ϵ ratio. The **X**-matrices were randomly split into training and test sets (3:1 ratio) 15 times (bootstrapping),¹¹ and each training set was regressed against the *y*-block of pIC_{50} values through partial least squares (PLS) to give the MIA-QSAR models.¹² The models were validated according to the best practices in QSAR.¹³ This workflow, as well as the building of the MIA-plots, was carried out using the Chemoface program.¹⁴

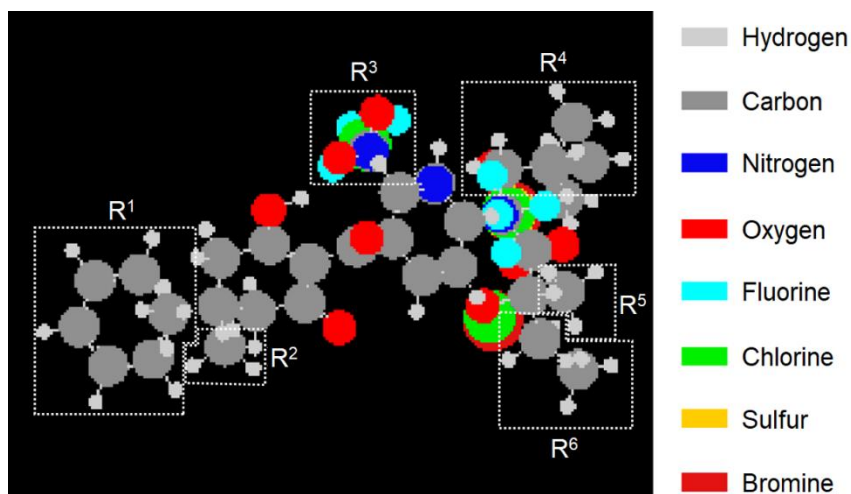


Figure 2. Superposed molecular images demonstrating the congruent scaffold and the variable moieties.

The molecular docking studies were carried out to validate the MIA-QSAR outcomes and aid the interpretation of ligand-enzyme interactions responsible for the bioactivities. The crystal structure of *A. thaliana* HPPD (*AtHPPD*, EC 1.13.11.27) complexed with the inhibitor 1,5-dimethyl-6-(2-oxidanyl-6-oxidanylidene-cyclohexen-1-yl)carbonyl-3-(3-trimethylsilylprop-2-ynyl)quinazoline-2,4-dione (**NJX**) and cobalt (II) ion was taken from the Protein Data Bank (PDB ID: 6M6D; resolution: 1.80 Å). Hydrogen atoms were added to the *AtHPPD* structure, and all the crystal water molecules (except H₂O644) were removed before modelling. The electrostatic charges on various atoms were defined by the AMBER force field. The structure was refined using the steepest descents and the conjugate gradients algorithms, until the final root-mean-square energy gradient was less than 0.10 kcal mol⁻¹. First, all the heavy atoms were constrained and only hydrogens were relaxed in a first step. In the second step, all the backbone atoms were constrained, and the side chains were relaxed. Finally, after removing all the constraints, the whole molecule was relaxed. These constraints were designed to remove the close contacts that may have been overlooked during the addition of hydrogens. All ligands were constructed and optimized with AMBER force field (GAFF) parameters and AM1-BCC charges. The protein and ligand structures were prepared with the Discovery Studio software. Docking calculations were performed on the two molecules using the Molegro Virtual Docker software,¹⁵ a fast and flexible docking program that gives the most likely conformation of ligand binding to a macromolecule. The Moldock score [GRID] function is a hybrid search algorithm that combines the differential evolution

optimization technique with a cavity prediction algorithm during the searching procedure, which allows fast and accurate recognition of potential binding modes, the poses. It is derived from the Piecewise Linear Potential (PLP), a simplified potential whose parameters are fit to protein-ligand structures and binding data scoring functions and further extended in GEMDOCK program (Generic Evolutionary Method for molecular DOCK) with a new hydrogen bonding term and new charge schemes.¹⁵ The grid-based scoring functions precalculated potential-energy values on an evenly spaced cubic grid to speed up calculations. The energy potential was evaluated using tri-linear interpolation between relevant grid points. A grid of 0.30 Å in resolution and 10 Å in radius from the inhibitor as the binding site was used. MolDock SE (Simplex Evolution) was used as a search algorithm, and the maximum number of poses to generate was 100. Potential binding sites were identified using the built-in cavity detection algorithm. A grid covering the protein was created and at every grid point a sphere was placed. It was checked whether this sphere overlaps with any of the spheres determined by the van der Waals radii of the protein atoms. The cavities that are identified by the cavity detection algorithm were then used by the guided differential evolution search algorithm to focus the search, to that specific area during the docking simulation.

3. Results and discussion

The first step to design novel triketone derivatives as HPPD inhibitors was to develop QSAR models with MIA descriptors encoding atomic van der Waals radii (r_{vdW}), Pauling's electronegativity (ϵ), and the r_{vdW}/ϵ ratio. To attest for the reliability and predictive capability of these models, the MIA descriptors for 15 randomly selected training sets were regressed against the pIC₅₀ values of the 68 triketones of Table 1 using partial least squares regression. The modeling performance of all three models were similar (Table 2), demonstrating that both r_{vdW} and ϵ parameters explain the data variance, that is, steric, hydrophobic, and electrostatic interactions play equally relevant roles in the biological environment. Moreover, the statistical metrics indicate the models were adjusted ($r^2 > 0.6$) but not overfitted (${}^c r_p^2 > 0.5$),¹⁶ and were both internally (leave-one-out cross-validation $q^2 > 0.5$)¹⁷ and externally (r_{pred}^2 and $r_m^2 > 0.5$)¹⁸ validated. Therefore, the regression parameters of these models were applied to predict the pIC₅₀ values of unprecedented triketone analogs.

Prior to using QSAR models to predict the inhibitory activity of new compounds, it is worth tracing tendencies by observing the experimental profile within the data set. Table 1 shows that compounds with nucleus **A** are generally more active than those with nucleus **B**, since the average pIC₅₀ are 6.45 and 5.59, respectively. The least active compound, **B11**,

possesses methyl groups as R¹ and R² substituents, unlike the most active compound **A41** and mesotrione, whose substituents are hydrogens. In addition, both **A41** and mesotrione contain a nitro group as R³, while **B11** has a 5-Br substituent; therefore, hydrophobic halogens at R³ appear to disfavor the biological activity. Thus, the proposition of successful candidates should be more assertive based on this observation. Moreover, MIA-plots allow for analysis of substituent effects on the pIC₅₀ values.

Table 1. The series of 68 triketones with scaffolds **A** and **B**, and the respective pIC₅₀ data (IC₅₀ in mol L⁻¹).^a

Compounds	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	pIC ₅₀
A1	H	H	Cl	CH ₃	H	CH ₃	6.47
A2	CH ₃	CH ₃	Cl	CH ₃	H	CH ₃	6.16
A3	H	H	Cl	C ₂ H ₅	H	CH ₃	6.44
A4	CH ₃	CH ₃	Cl	C ₂ H ₅	H	CH ₃	6.34
A5	H	H	Cl	allyl	H	CH ₃	6.58
A6	CH ₃	CH ₃	Cl	allyl	H	CH ₃	6.32
A7	H	H	Cl	propargyl	H	CH ₃	6.64
A8	CH ₃	CH ₃	Cl	propargyl	H	CH ₃	6.51
A9	H	H	Cl	CH ₃	H	C ₂ H ₅	6.51
A10	CH ₃	CH ₃	Cl	CH ₃	H	C ₂ H ₅	6.23
A11	H	H	Cl	C ₂ H ₅	H	C ₂ H ₅	6.59
A12	H	H	Cl	propargyl	H	C ₂ H ₅	6.63
A13	H	H	Cl	CH ₃	CH ₃	CH ₃	6.34
A14	H	H	Cl	C ₂ H ₅	CH ₃	CH ₃	6.58
A15	CH ₃	CH ₃	Cl	C ₂ H ₅	CH ₃	CH ₃	5.68
A16	H	H	Cl	allyl	CH ₃	CH ₃	6.54
A17	CH ₃	CH ₃	Cl	allyl	CH ₃	CH ₃	5.62
A18	H	H	Cl	propargyl	CH ₃	CH ₃	6.28
A19	H	H	NO ₂	CH ₃	H	CH ₃	6.57
A20	CH ₃	CH ₃	NO ₂	CH ₃	H	CH ₃	6.51
A21	H	H	NO ₂	C ₂ H ₅	H	CH ₃	6.44
A22	CH ₃	CH ₃	NO ₂	C ₂ H ₅	H	CH ₃	6.58
A23	H	H	NO ₂	allyl	H	CH ₃	6.50
A24	CH ₃	CH ₃	NO ₂	allyl	H	CH ₃	6.49
A25	H	H	NO ₂	propargyl	H	CH ₃	6.49
A26	CH ₃	CH ₃	NO ₂	propargyl	H	CH ₃	6.57
A27	H	H	NO ₂	CH ₃	H	C ₂ H ₅	6.49
A28	CH ₃	CH ₃	NO ₂	CH ₃	H	C ₂ H ₅	6.38
A29	H	H	NO ₂	C ₂ H ₅	H	C ₂ H ₅	6.61
A30	CH ₃	CH ₃	NO ₂	C ₂ H ₅	H	C ₂ H ₅	6.67

A31	H	H	NO ₂	allyl	H	C ₂ H ₅	6.66
A32	CH ₃	CH ₃	NO ₂	allyl	H	C ₂ H ₅	6.57
A33	H	H	NO ₂	propargyl	H	C ₂ H ₅	6.56
A34	CH ₃	CH ₃	NO ₂	propargyl	H	C ₂ H ₅	6.67
A35	H	H	NO ₂	CH ₃	CH ₃	CH ₃	6.23
A36	CH ₃	CH ₃	NO ₂	CH ₃	CH ₃	CH ₃	6.01
A37	H	H	NO ₂	C ₂ H ₅	CH ₃	CH ₃	6.38
A38	CH ₃	CH ₃	NO ₂	C ₂ H ₅	CH ₃	CH ₃	6.27
A39	H	H	NO ₂	allyl	CH ₃	CH ₃	6.76
A40	CH ₃	CH ₃	NO ₂	allyl	CH ₃	CH ₃	6.54
A41	H	H	NO ₂	propargyl	CH ₃	CH ₃	6.81
A42	CH ₃	CH ₃	NO ₂	propargyl	CH ₃	CH ₃	6.65
B1	H	H	6-Cl				6.58
B2	CH ₃	H	6-Cl				5.51
B3	CH ₃	CH ₃	6-Cl				5.48
B4	Ph	H	6-Cl				5.35
B5	H	H	2-CF ₃				5.65
B6	CH ₃	H	2-CF ₃				5.46
B7	CH ₃	CH ₃	2-CF ₃				5.36
B8	Ph	H	2-CF ₃				4.98
B9	H	H	5-Br				5.57
B10	CH ₃	H	5-Br				5.36
B11	CH ₃	CH ₃	5-Br				4.86
B12	H	H	6-F				5.46
B13	CH ₃	H	6-F				5.06
B14	Ph	H	6-F				5.72
B15	H	H	6-CF ₃				6.15
B16	CH ₃	H	6-CF ₃				6.10
B17	CH ₃	CH ₃	6-CF ₃				5.40
B18	Ph	H	6-CF ₃				6.02
B19	H	H	6-Br				5.76
B20	CH ₃	H	6-Br				5.56
B21	CH ₃	CH ₃	6-Br				4.95

B22	Ph	H	6-Br	6.04
B23	H	H	5-Cl, 6-Cl	5.99
B24	CH ₃	H	5-Cl, 6-Cl	5.79
B25	CH ₃	CH ₃	5-Cl, 6-Cl	5.71
Mesotrione				6.60

^a Where applicable, the stereogenic carbon attached to R¹ and R² was *R* configuration, while the stereogenic carbon attached to R⁵ and R⁶ was *S* configuration.

Table 2. Average statistical metrics of calibration and validation of the MIA-QSAR models (mean for 15 cycles).

Parameter	r _{vdW}	ε	r _{vdW} /ε
PLS components	2.7 ± 0.7	2.2 ± 0.4	3.3 ± 0.8
RMSEC	0.226 ± 0.017	0.230 ± 0.015	0.220 ± 0.015
r ²	0.813 ± 0.027	0.805 ± 0.023	0.823 ± 0.022
RMSE _{y-rand}	0.452 ± 0.021	0.456 ± 0.021	0.450 ± 0.020
r ² _{y-rand}	0.252 ± 0.048	0.229 ± 0.037	0.260 ± 0.045
^c r ² _p	0.675 ± 0.035	0.680 ± 0.029	0.680 ± 0.029
RMSECV	0.295 ± 0.016	0.294 ± 0.013	0.295 ± 0.016
q ²	0.685 ± 0.028	0.681 ± 0.024	0.685 ± 0.027
RMSEP	0.294 ± 0.049	0.287 ± 0.048	0.291 ± 0.048
r ² _{pred}	0.684 ± 0.091	0.709 ± 0.076	0.702 ± 0.099
r ² _m	0.523 ± 0.129	0.526 ± 0.111	0.566 ± 0.139

To guide the proposition of new agrochemical candidates, the MIA plots of variable importance in projection (VIP) scores and PLS regression coefficients (**b**) were analyzed (Figure 3). These heat contour maps indicate, respectively, how much and how (increasing or decreasing pIC₅₀) the substituents R¹-R⁶ affect the biological response. The influence of each substituent increases as the atom color ranges from dark blue to dark red in the VIP plots. Accordingly, it has been found that substituents in all positions, excepting R² and R⁴, have significant effect on the pIC₅₀ values. In turn, the **b** plots show that the CH₃ group in R¹ and R⁵, and **A**-Cl, **B**-Br, and **B**-2-CF₃ groups in R³ are deleterious to the inhibition. However, electronegative groups in R³, namely **B**-6-F, **B**-6-CF₃, and **A**-NO₂, enhance the pIC₅₀ values. Substituents **B**-6-Cl in R³, and CH₃ or Et in R⁶ have no or slightly positive effects on inhibition. In addition to this guidance, the combination of substituents attached to the most

active compounds and mesotrione is a rational strategy to generate new derivatives. Based on these approaches, a total of 34 new triketone analogs were designed and the corresponding pIC_{50} values were estimated using the MIA-QSAR models (Table 3).

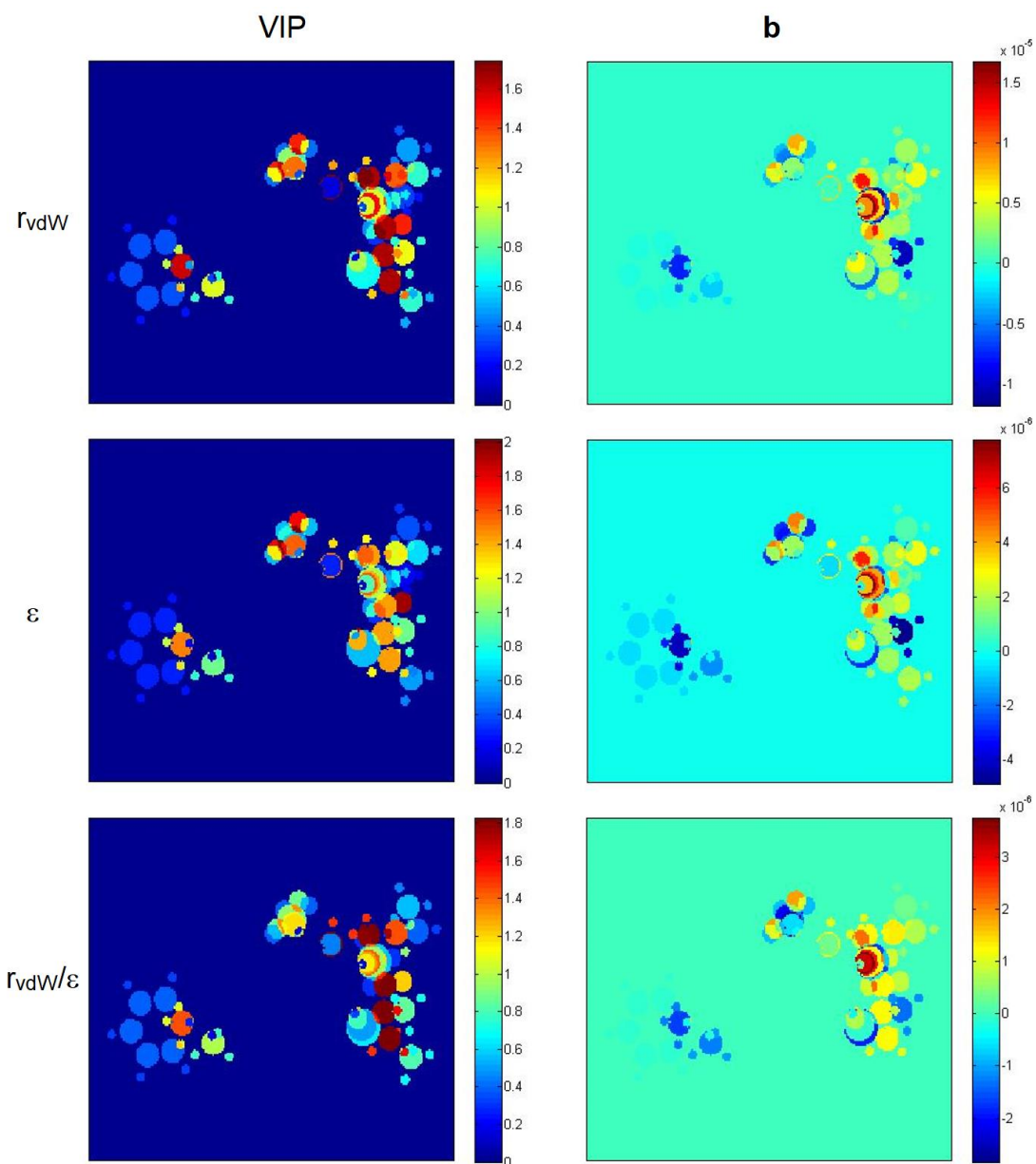


Figure 3. MIA plots for the 68 triketones analyzed using PLS-MIA-QSAR.

Table 3. Proposed triketones with nuclei **A (P1-P21)** and **B (P22-P34)**, and respective pIC_{50} obtained from the MIA-QSAR models.

Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	r _{vdW}	ε	r _{vdW} /ε	Avg.
P1	CH ₃	CH ₃	Cl	propargyl	CH ₃	CH ₃	6.15	6.14	6.14	6.14
P2	CH ₃	CH ₃	NO ₂	propargyl	CH ₃	C ₂ H ₅	6.40	6.41	6.36	6.39
P3	CH ₃	H	NO ₂	propargyl	H	C ₂ H ₅	6.65	6.64	6.64	6.64
P4	CH ₃	H	NO ₂	propargyl	CH ₃	CH ₃	6.46	6.48	6.44	6.46
P5	H	H	CF ₃	propargyl	CH ₃	CH ₃	6.44	6.37	6.47	6.43
P6	H	H	CF ₃	propargyl	H	C ₂ H ₅	6.63	6.52	6.66	6.61
P7	H	H	CF ₃	allyl	CH ₃	CH ₃	6.40	6.32	6.43	6.38
P8	H	H	CF ₃	allyl	CH ₃	C ₂ H ₅	6.47	6.39	6.47	6.45
P9	H	H	NO ₂	propargyl	CH ₃	C ₂ H ₅	6.75	6.77	6.69	6.74
P10	CH ₃	H	NO ₂	propargyl	CH ₃	C ₂ H ₅	6.54	6.56	6.48	6.52
P11	H	CH ₃	Cl	propargyl	CH ₃	CH ₃	6.36	6.35	6.35	6.35
P12	CH ₃	H	Cl	propargyl	CH ₃	CH ₃	6.28	6.29	6.26	6.28
P13	H	CH ₃	NO ₂	propargyl	H	CH ₃	6.65	6.62	6.70	6.66
P14	CH ₃	H	NO ₂	propargyl	H	CH ₃	6.58	6.56	6.60	6.58
P15	H	CH ₃	NO ₂	propargyl	H	C ₂ H ₅	6.72	6.70	6.74	6.72
P16	H	CH ₃	Cl	allyl	CH ₃	CH ₃	6.31	6.30	6.32	6.31
P17	CH ₃	H	Cl	allyl	CH ₃	CH ₃	6.24	6.24	6.22	6.23
P18	H	CH ₃	NO ₂	allyl	H	CH ₃	6.60	6.58	6.66	6.61
P19	CH ₃	H	NO ₂	allyl	H	CH ₃	6.53	6.51	6.57	6.54
P20	H	CH ₃	NO ₂	allyl	H	C ₂ H ₅	6.68	6.65	6.70	6.68
P21	CH ₃	H	NO ₂	allyl	H	C ₂ H ₅	6.61	6.59	6.61	6.60
P22	H	H	2-Cl, 6-Cl				5.87	5.90	5.83	5.87
P23	H	H	2-NO ₂ , 6-Cl				6.05	6.10	6.02	6.05
P24	H	H	6-SO ₂ CH ₃				6.17	6.10	6.28	6.18
P25	H	H	6-SO ₂ CH ₃ , 2-NO ₂				6.35	6.36	6.42	6.38
P26	H	H	6-SO ₂ CH ₃ , 2-Cl				6.17	6.17	6.24	6.19
P27	H	H	6-SO ₂ CH ₃ , 2-CF ₃				6.12	6.04	6.23	6.13
P28	H	H	2-Cl				5.58	5.69	5.52	5.60
P29	H	H	2-NO ₂				5.77	5.88	5.71	5.78
P30	CH ₃	H	2-Cl				5.38	5.48	5.31	5.39

P31	CH ₃	H	2-NO ₂	5.56	5.67	5.49	5.57
P32	CH ₃	H	6-SO ₂ CH ₃ , 2-NO ₂	6.14	6.15	6.21	6.17
P33	CH ₃	H	6-SO ₂ CH ₃ , 2-NO ₂	6.18	6.19	6.28	6.22
P34	CH ₃	CH ₃	6-SO ₂ CH ₃ , 2-NO ₂	6.05	6.04	6.16	6.08

According to the predicted inhibition data obtained with the MIA-QSAR models, 8 out of 17 compounds with higher pIC₅₀ than mesotrione are newly proposed molecules, being **P9** the third most active (Figure 4). A proposed synthetic route to obtain **P9** is given in Scheme 1. The predicted pIC₅₀ values are practically insensitive to the stereogenic carbon where R¹ and R² are attached (for instance, compare the enantiomers **P11** and **P12**). Nevertheless, in an ideal scenario, a compound should firstly be absorbed by the plant and then permeate the cell to reach the biological target, besides to be environmentally friendly. Considering that mesotrione can be applied post-emergence, more lipophilic analogs are expected to be better absorbed by leaves. Also, because lipophilic compounds have low mobility in soils rich in organic matter, they are less prone to leach out and reach the groundwater before decaying. Organic-carbon partition coefficient (K_{OC}) of a compound is related to its lipophilicity¹⁹ and, therefore, may be used to evaluate the risk to contaminating the groundwater; according to the EPA, chemicals with $K_{OC} < 1,000 \text{ mL g}^{-1}$ are considered moderately to highly mobile in the soil. In turn, the European standard for the distribution coefficient (K_d) is $< 10 \text{ mL g}^{-1}$.²⁰ Therefore, the octanol-water partition coefficient ($\log P$) can be a proper parameter for bioactivity and environmental risk assessment. The calculated $\log P$ for the compounds studied herein are remarkably higher than that of mesotrione, whereas the most lipophilic compound was found to be **P6**, followed by **A30**, **P20**, and **P21** (Figure 4). It is worth mentioning, however, that high lipophilicity (particularly, $\log P$ ranging between 5 and 8)²¹ may cause other ecological issues, such as bioaccumulation.

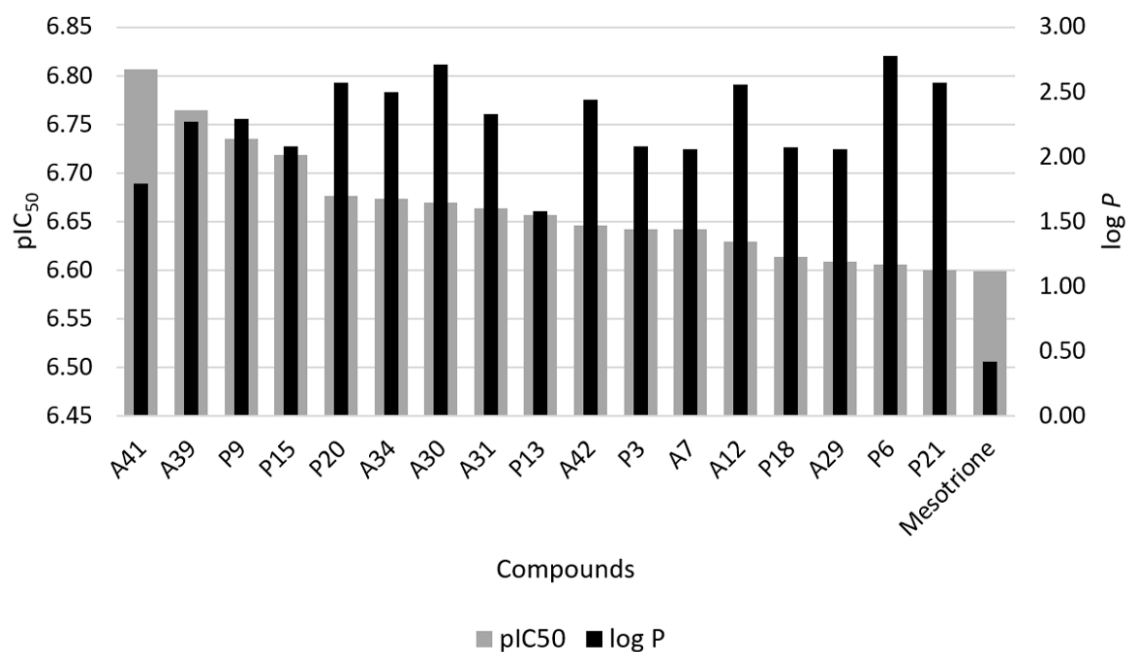
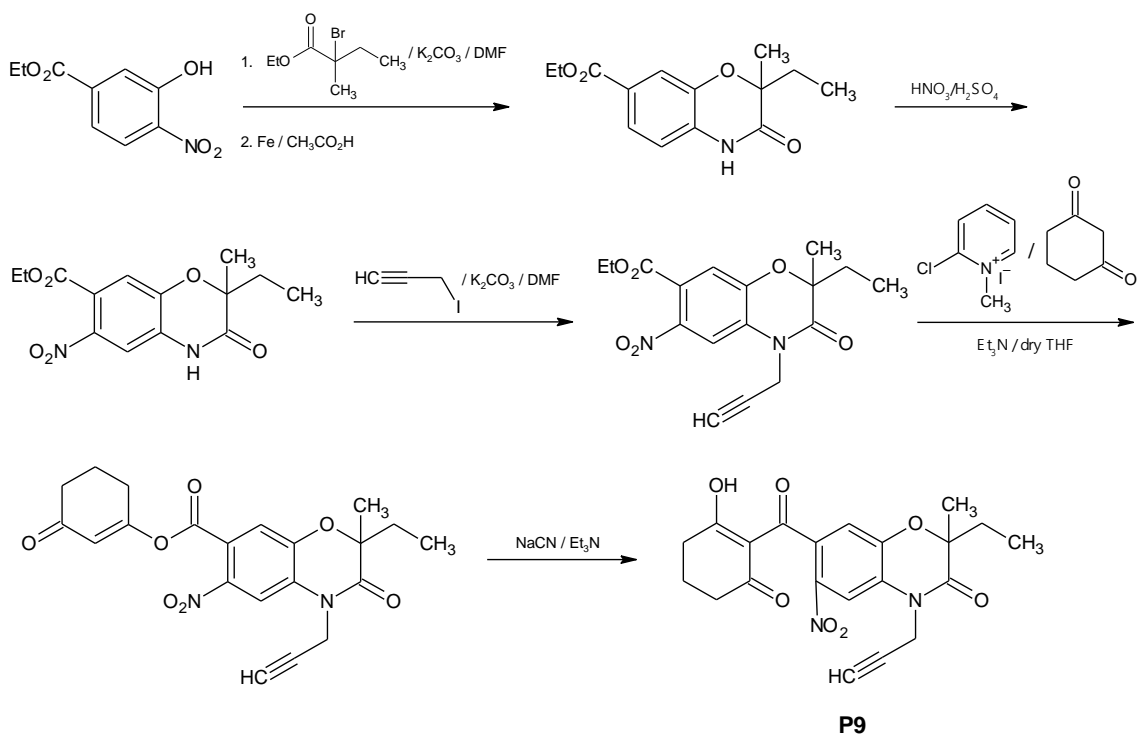


Figure 4. Calculated pIC₅₀ and log *P* for compounds with pIC₅₀ higher than the mesotrione.



Scheme 1. Possible synthetic route to obtain compound **P9**, in agreement to a similar procedure described in the literature.⁵

To further explore the interactions that drive the inhibitory activity of the compounds toward the HPPD enzyme, docking studies were carried out for the mesotrione, **NJX** (the

native ligand), library triketones, and also for the proposed analogs, with special emphasis to the most promising candidates.

The *Arabidopsis thaliana* HPPD (*AtHPPD*) binding site analyzed was defined as a subset region next to His205, His287, Gln272, Gln286, and the metal. A cavity of 263.68 Å³ was selected. To account for side chain flexibility, the amino acid residues around 6.0 Å from center of cavity were selected to be kept flexible during the docking simulation. The docking studies were validated by redocking the **NJX** into the cavity and superimposing it onto co-crystallized complex. Two initial conformations were considered: the original from the Protein Data Bank (**NJX-1**) and the energy minimized conformation (**NJX-2**). Docking programs that are able to predict poses with RMSD lower than 1.5 to 2.0 Å (depending on ligand size) are considered successful.²² The poses (**NJX-1** and **NJX-2**) were obtained, and the protein-pose complex structures were superimposed to the crystal structure. RMSD's of 0.41 Å and 0.49 Å were observed for **NJX-1** and **NJX-2**, respectively. The same method previously applied was used in the docking process. The poses were selected according to two criteria: the orientation of **NJX** in the binding site (experimental information) and the pose with lowest energy. The standard drug mesotrione binds at the active site and mediates several interactions including ionic interactions with Co²⁺ ion; hydrogen bonding with H₂O644, the side chains of His226 and His308; π - π stacking interaction with Phe381 and Phe424; π -sulfur interaction with Phe424 (Figure 5). **NJX** interacts with the Co²⁺ ion, H₂O644, His226, His308, Met335, Glu394, Phe381, and Phe424. The binding mode of the most active compound (**A41**) depicts that the substituent at R¹ position, next to Leu368 and Phe381, is favorable to perform hydrophobic interactions; the hydrophobic aromatic feature on the ring is favorable to π - π stacking interactions with Phe381 and Phe424; the triketone moiety interacts with Co²⁺ ion, H₂O644, and His226; the NO₂ group interacts with Gln307. On the other hand, the least active compound (**B11**) shows a steric hindrance between R¹/R² substituents (methyl groups) and Ser267. Interestingly, all compounds with two methyl groups (R¹ and R²) are less potent than the compounds with one methyl or hydrogen substituent. The proposed compound **P9** was docked into the *AtHPPD* binding site and the interactions with the amino acid residues are shown in Figure 5.

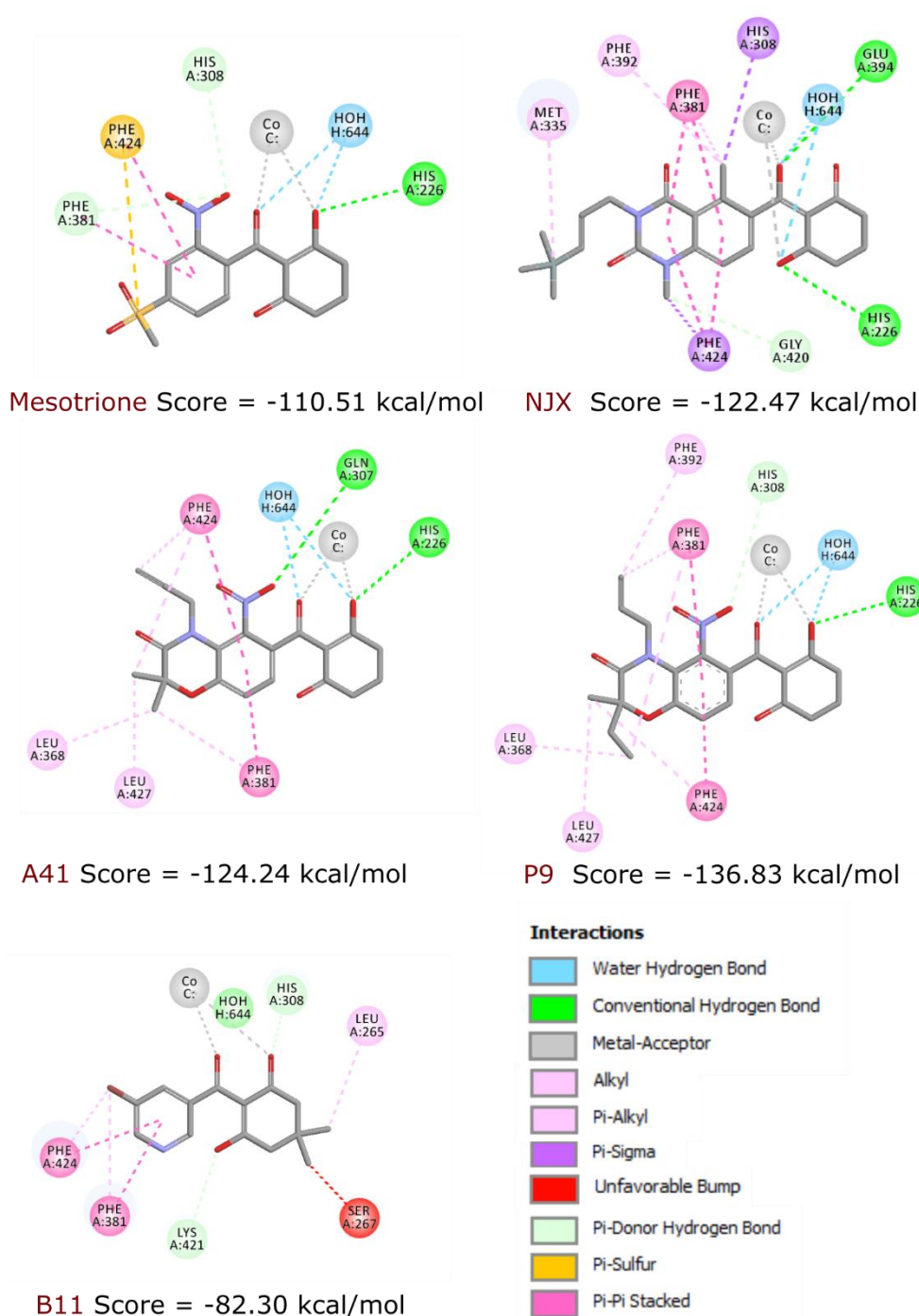


Figure 5. Important ligand-enzyme interactions obtained from the docking studies.

The performance of the docking models is mostly measured in terms of their ability to discriminate between active and inactive compounds by calculating the docking energies. The energy calculations by MolDock score of each pose were plotted against pIC_{50} (Figure 6), giving a determination coefficient of 0.61, which means that a good correlation between the experimental and calculated data was obtained ($R^2 > 0.5$). Finally, the outcomes from the docking hint indicate that the proposed compound **P9** forms a more stable complex (-136.8

kcal mol⁻¹) than mesotrione (-110.5 kcal mol⁻¹), **NJX** (-122.5 kcal mol⁻¹), and **P41** (-124.2 kcal mol⁻¹). The *R* enantiomer of **P9** was also evaluated and demonstrated to be 8 kcal mol⁻¹ less stable than the *S*, due to a more effective interaction with the metal in the latter. This implies that our candidate compound **P9** possesses high affinity to the *At*HPPD active site, and it is a promising herbicide candidate.

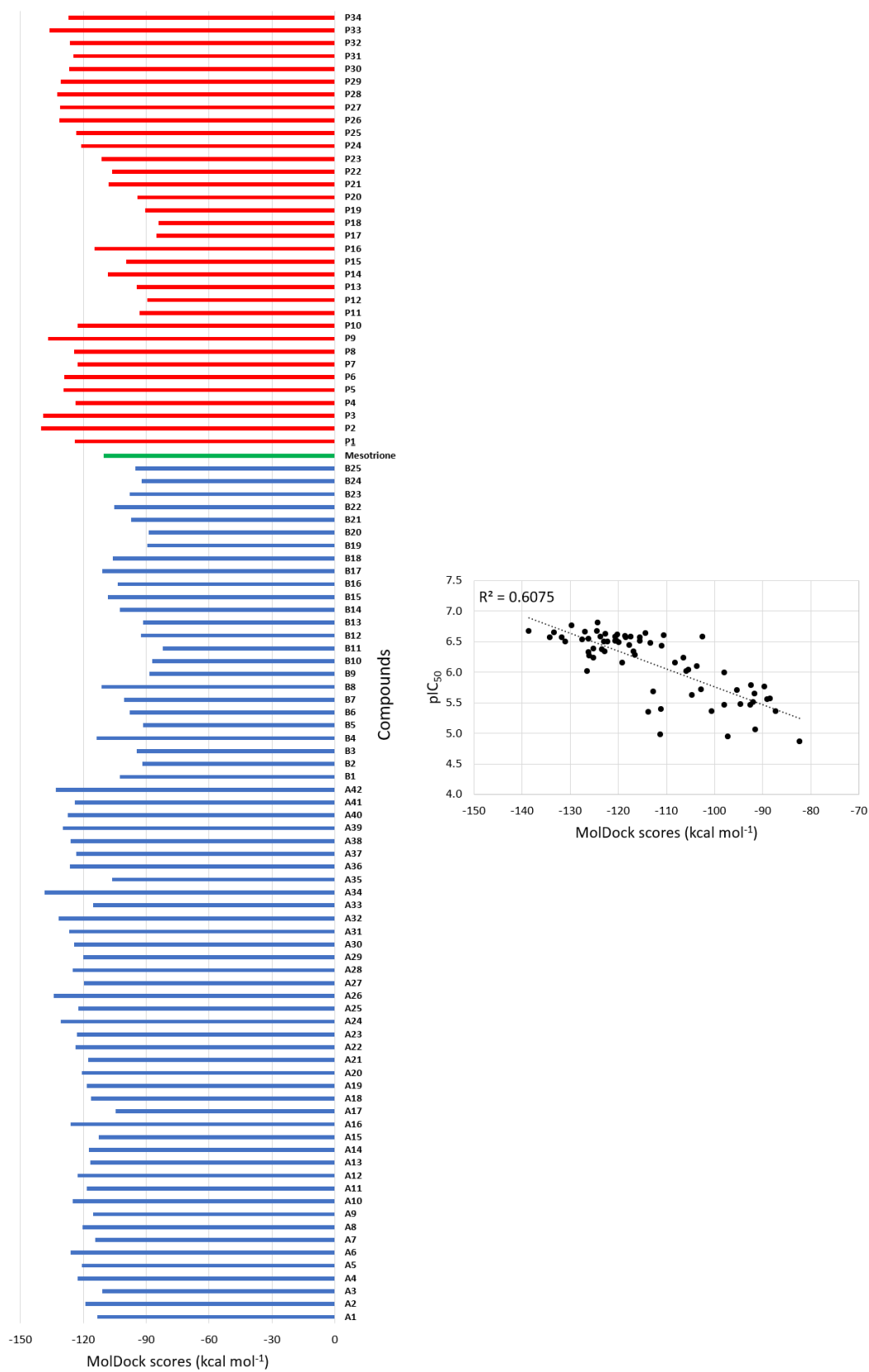


Figure 6. Docking scores (kcal mol⁻¹) for the mesotrione (green), library (blue) and proposed (red) compounds, and the correlation plot of docking scores *versus* experimental pIC₅₀.

4. Conclusion

A total of 8 triketone derivatives were proposed with predicted HPPD activity higher than of the commercial mesotrione, being **P9** the most relevant proposal and also a highly lipophilic compound. This feature is an important parameter for environmental risk assessment related to runoff, but also for the absorption by plant leaves. The docking analyses demonstrated that **P9** forms a more stable complex with the *Az*HPPD enzyme than compound **A41**, which is the best ranked inhibitor according to the experimental values of pIC_{50} . Although chemical syntheses and bioassays are indeed required to confirm the theoretical evidence, our findings provided guidelines to design new chemicals with substructure **A**, which are likely to act as herbicides.

Acknowledgements

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, funding code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant number 306830/2021-3), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

References

1. Penfold AR, The Essential Oil of *Melaleuca linariifolia* (Smith) and *M. alternifolia* (Cheel). *Pamphlet Collection* **59**:306-324 (1926).
2. Fu Y, Zhang SQ, Liu YX, Wang JY, Gao S, Zhao LX and Ye F, Design, synthesis, SAR and molecular docking of novel green niacin-triketone HPPD inhibitor. *Ind. Crops Prod.* **137**:566-575 (2019).
3. Moran GR, 4-Hydroxyphenylpyruvate dioxygenase. *Arch. Biochem. Biophys.* **433**:117-128 (2005).
4. Küpper A, Peter F, Zollner P, Lorentz L, Tranel PJ, Beffa R and Gaines TA, Tembotrione detoxification in 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitor-resistant Palmer amaranth (*Amaranthus palmeri* S. Wats.). *Pest Manag. Sci.* **74**:2325-2334 (2018).
5. Li HB, Li L, Li JX, Han TF, He JL and Zhu YQ, Novel HPPD inhibitors: triketone 2H-benzo[b][1,4]oxazin-3(4H)-one analogs. *Pest Manag. Sci.* **74**:579-589 (2018).
6. Freitas MR, Barigye SJ and Freitas MP, Coloured chemical image-based models for the prediction of soil sorption of herbicides. *RSC Adv.* **5**:7547-7553 (2015).
7. Barigye SJ, Duarte MH, Nunes CA and Freitas MP, MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. *RSC Adv.* **6**:49604-49612 (2016).

8. Chahal PS, Jugulam M and Jhala AJ, Basis of Atrazine and Mesotrione Synergism for Controlling Atrazine- and HPPD Inhibitor-Resistant Palmer Amaranth. *Agron. J.* **111**:3265-3273 (2019).
9. Jacobs KE, Butts-Wilmsmeyer CJ, Ma R, O'Brien SR and Riechers DE, Association between metabolic resistances to atrazine and mesotrione in a multiple-resistant waterhemp (*Amaranthus tuberculatus*) population. *Weed Sci.* **68**:358-366 (2020).
10. Dennington RD, Keith TA and Millam MJ, GaussView 5.0 (2008).
11. Kiralj R and Ferreira MMC, Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.* **20**:770-787 (2009).
12. Geladi P and Kowalski BR, Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**:1-17 (1986).
13. Tropsha A, Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **29**:476-488 (2010).
14. Nunes CA, Freitas MP, Pinheiro ACM and Bastos SC, Chemoface: a novel free user-friendly interface for chemometrics. *J. Braz. Chem. Soc.* **23**:2003-2010 (2012).
15. Thomsen R and Christensen MH, MolDock: A New Technique for High-Accuracy Molecular Docking. *J. Med. Chem.* **49**:3315-3321 (2006).
16. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S and Das RN, 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 (2013).
17. Consonni V, Ballabio D and Todeschini R, Comments on the Definition of the Q2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **49**:1669-1678 (2009).
18. Mitra I, Saha A and Roy K, Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol. Simul.* **36**:1067-1079 (2010).
19. Sabljic A, Güsten H, Verhaar H, Hermens J, QSAR modelling of soil sorption. Improvements and systematics of logK_{OC} versus logK_{OW} correlations. *Chemosphere* **31**:4489-4514 (1995).
20. Park WP, Chang KM, Hyun HN, Boo KH, Koo BJ, Sorption and leaching characteristics of pesticides in volcanic ash soils of Jeju Island, Korea. *Appl. Biol. Chem.* **63**:71 (2020).

21. Lecloux A, Libre, J-M, Lyne R, ten Berge W, Thompson R, Vandepitte V, Weiss M, Holt M, Persistent Organic Pollutants (POPs) Response to UNEP/INC/CEG-I Annex 1, Doc. N° 41, ECETOC, Brussels (2000).
22. Pagadala NS, Syed K and Tuszynski J, Software for molecular docking: a review. *Biophys. Rev.* **9**:91-102 (2017).

Molecular Modeling Study of Two Combined Series of Triketones with HPPD Inhibitory Activity

Luiz R. Capucho,^a Elaine F. F. da Cunha,^a Matheus P. Freitas^{a,*}

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

* Corresponding author: matheus@ufla.br

Abstract

Triketones are suitable compounds to inhibit the 4-hydroxyphenylpyruvate dioxygenase (HPPD) and an important strategy to eliminate agricultural weeds. We report herein the quantitative structure-activity relationship (QSAR) modeling and docking studies for a series of triketone-quinoline hybrids and 2-(aryloxyacetyl)cyclohexane-1,3-diones with the aim of proposing new chemical candidates with improved performance as herbicides. The QSAR models obtained were reliable and predictive (average r^2 , q^2 , and r^2_{pred} of 0.72, 0.51, and 0.71, respectively). Guided by multivariate image analysis of the PLS regression coefficients and variable importance in projection scores, the substituent effects could be analyzed and a promising derivative with $R^1 = \text{H}$, $R^2 = \text{CN}$, and $R^3 = 5,7,8\text{-triCl}$ at the triketone-quinoline scaffold (**P18**) was proposed. Docking studies demonstrated that π - π stacking interactions and specific interactions between the substituents and amino acid residues in the binding site of the *Arabidopsis thaliana* HPPD (*AtHPPD*) enzyme handle the desired bioactivity.

Keywords: Triketones; HPPD enzyme; QSAR; docking

Introduction

Triketone herbicides, remarkably mesotrione (Figure 1a), which was first marketed in 2001, are a relatively new class of 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitors used in agriculture (since 1977), especially in maize crops (Beaudegnies et al. 2009). These herbicides derive from leptospermone (Figure 1b), a natural herbicide which inhibits the biosynthesis of tocopherols and plastoquinone essential to carotenoid production; this leads to bleaching of leaves as chlorophyll is degraded, followed by plant death (Moran 2005).

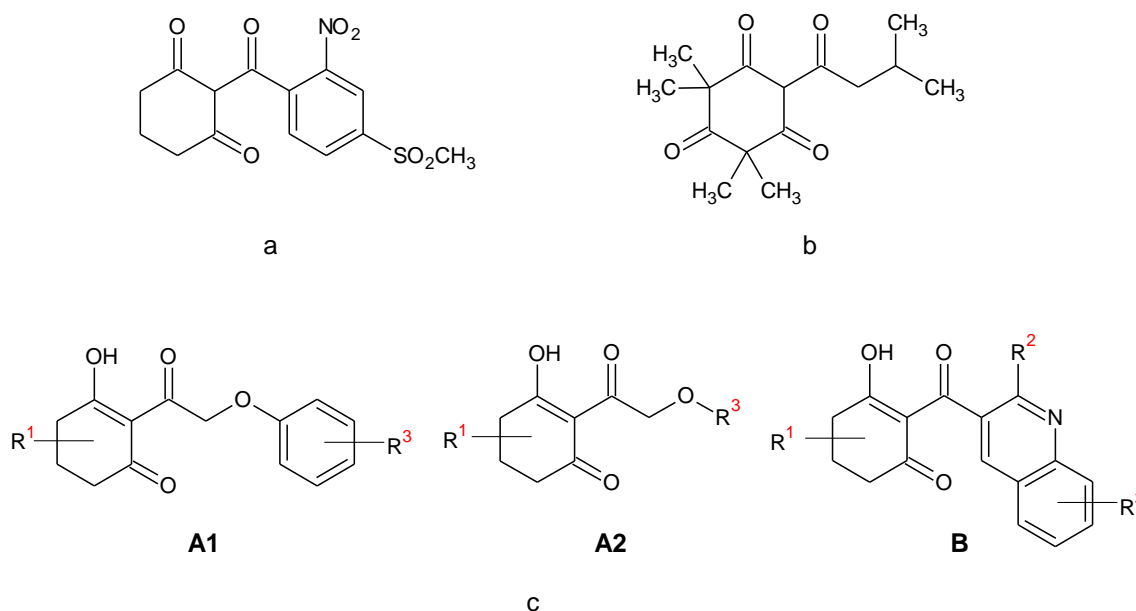


Figure 1. (a) Mesotrione, (b) leptospermone, and (c) triketones type **A1**, **A2**, and **B** studied herein.

Because the resistance of weeds to herbicides is an ongoing problem to be tackled in agriculture, new agrochemicals should be continuously developed. For example, the resistance to mesotrione in an *Amaranthus tuberculatus* population has been attributed to genetics and inheritance related to a non-target-site based mechanism (Kaundun et al. 2017; Oliveira et al. 2018). Accordingly, two congeneric series of triketones have been synthesized and tested against *Arabidopsis thaliana* HPPD (*AtHPPD*), demonstrating a similar to better performance compared to mesotrione (Figure 1c) (Wang et al. 2015; 2016). According to the authors, coordination with Fe^{II} and π - π stacking interactions in the binding site provide high affinity to compounds. In addition, with a few exceptions, electron-withdrawing groups attached to the scaffold (such as chlorine) produced enhanced effects to the bioactivity.

Considering these findings, both series of compounds can be joined into a single data set to obtain a large library of molecules for the modeling of the bioactivities using quantitative structure-activity relationship (QSAR) analysis. Subsequently, new derivatives can be designed and proposed as possible candidates for HPPD inhibition.

Our focus herein was to model the *At*HPPD inhibition constants, described in terms of pK_i , of 97 triketones using multivariate image analysis applied to QSAR (MIA-QSAR) (Freitas et al. 2015). This procedure allows for the analysis of the effects caused by substituent groups to the bioactivities through the MIA-plots, which are contour maps that describe how and how much certain molecular moieties affect the pK_i values (Barigye et al. 2016). In addition, the binding affinities responsible for the biological activities were interpreted using docking studies. Finally, since lipophilicity may affect both the bioactivity and environment, the octanol-water partition coefficients ($\log P$) were calculated for the library and proposed compounds.

Material and Methods

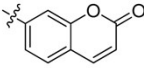
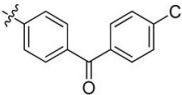
Two series of triketone molecules composed of 38 triketone-quinoline hybrids (Wang et al. 2015) and 58 2-(aryloxyacetyl)cyclohexane-1,3-diones (Wang et al. 2016), and also mesotrione were joined to compose an enlarged data set of 97 triketone derivatives with known *At*HPPD inhibition K_i (in μM) (Table 1). This allows for molecular diversity to propose novel herbicide candidates. For the MIA-QSAR modeling, whose descriptors derive from RGB pixels of molecular images that encode atomic properties and substituent orientations, the 97 molecules were sketched as ball-and-stick structures in the GaussView program (Dennington et al., 2008), keeping the congruent moiety (the common scaffold) aligned, in order to assess the variance in the pK_i values as a function of the variable chemical groups. Figure 2 depicts the superposition of the 97 chemical images to illustrate the congeneric and variable moieties of the molecules. The images of 204×236 pixels each were unfolded to row vectors of $1 \times 48,144$ pixels and then grouped to obtain an \mathbf{X} -matrix of $97 \times 48,144$ pixels. The color for each atom in the chemical image, represented as an RGB digit (from 0 to 765), was replaced with the corresponding van der Waals radius (r_{vdW}), electronegativity (ϵ), and r_{vdW}/ϵ ratio, giving three data matrices for subsequent multivariate analysis. The calibration models were generated by regressing each \mathbf{X} against the \mathbf{y} -block containing the pK_i values through partial least squares (PLS) regression (Geladi and Kowalski

1986). Validation of the experiments was performed using **y**-randomization tests (10 cycles) (Mitra et al. 2010), leave-one-out cross-validation (q^2) (Consonni et al. 2009), and external validation for 25% of a randomly selected test set of compounds (r^2_{pred} , avg. r^2_{m} , Δr^2_{m} , and CCC) (Roy et al. 2013; Chirico and Gramatica 2011). The PLS models were used to predict the $\text{p}K_{\text{i}}$ values of unprecedented derivatives originated from the combination of relevant substituents attached to the congruent scaffold. To analyze the effects of these substituents, MIA-plots based on PLS regression coefficients (**b**) and variable importance in projection (VIP) scores were built, giving detailed information on how (increasing or decreasing $\text{p}K_{\text{i}}$) and how much the substituents affect the herbicidal activity, respectively. These calculations were all performed using the Chemoface program (Nunes et al. 2012). The log P values for the library and proposed compounds were calculated for environmental risk assessment through the molinspiration.com.

Table 1. Substituted triketones with **A1**, **A2**, and **B** scaffolds and the respective $\text{p}K_{\text{i}}$ values (K_{i} in mol L⁻¹).

Cpd	R¹	R³	p<i>K</i>_i
A1-1	H	H	5.91
A1-2	5,5-diCH ₃	H	5.25
A1-3	H	2-CH ₃	6.05
A1-4	5,5-diCH ₃	2-CH ₃	5.80
A1-5^a	H	3-CH ₃	6.21
A1-6	H	4-CH ₃	6.01
A1-7	H	4-OCH ₃	6.52
A1-8	H	2-SCH ₃	6.22
A1-9	H	2-Cl	5.94
A1-10	5,5-diCH ₃	2-Cl	5.67
A1-11	H	3-Cl	7.12
A1-12^a	H	4-Cl	7.54
A1-13	H	2-CF ₃	6.87
A1-14	H	3-CF ₃	6.25
A1-15	H	4-CF ₃	5.98
A1-16^a	H	2-NO ₂	7.35
A1-17^a	H	4-SO ₂ CH ₃	6.45
A1-18	5-CH ₃	4-SO ₂ CH ₃	6.26

A1-19^a	H	2,3-diCl	6.95
A1-20^a	H	2,4-diCl	7.37
A1-21^a	H	2,5-diCl	6.92
A1-22	H	2,6-diCl	7.35
A1-23	H	3,4-diCl	6.57
A1-24^a	H	3,5-diCl	6.02
A1-25	5-CH ₃	2,4-diCl	6.71
A1-26	5,5-diCH ₃	2,4-diCl	5.32
A1-27^a	4,4-diCH ₃	2,4-diCl	6.62
A1-28	H	2,4-diBr	7.18
A1-29	5,5-diCH ₃	2,4-diBr	7.06
A1-30	H	2-CH ₃ -4-F	6.68
A1-31	H	2-CH ₃ -4-Cl	7.51
A1-32	5-CH ₃	2-CH ₃ -4-Cl	7.43
A1-33	5,5-diCH ₃	2-CH ₃ -4-Cl	6.96
A1-34	4,4-diCH ₃	2-CH ₃ -4-Cl	5.70
A1-35	H	2-CH ₃ -4-Br	7.11
A1-36	H	2-CH ₃ -4-NO ₂	7.11
A1-37	H	2-Cl-4-F	6.26
A1-38	H	2-Cl-4-NO ₂	7.47
A1-39	H	2-F-4-Cl	6.45
A1-40	H	2-NO ₂ -3-CH ₃	7.16
A1-41	H	3,5-diF-4-CN	7.09
A1-42	H	2,4,6-tri-Cl	6.74
A1-43	H	2,3,4,5,6-5F	7.26
A2-1	H	1-naphthyl	6.67
A2-2^a	5-CH ₃	1-naphthyl	5.89
A2-3	5,5-diCH ₃	1-naphthyl	5.45
A2-4	4,4-diCH ₃	1-naphthyl	5.95
A2-5	H	2-naphthyl	6.60
A2-6	5-CH ₃	2-naphthyl	5.44
A2-7^a	5,5-diCH ₃	2-naphthyl	5.44
A2-8	4,4-diCH ₃	2-naphthyl	7.10

A2-9^a	H		6-CN-2-naphthyl	6.09
A2-10^a	5,5-diCH ₃		6-CN-2-naphthyl	5.63
A2-11	H		6-Br-2-naphthyl	5.55
A2-12	5,5-diCH ₃		6-Br-2-naphthyl	5.41
A2-13	H		1-Br-naphthalen-2-yl	6.72
A2-14	H			5.50
A2-15	H			6.43
B-1	H	H	H	7.12
B-2^a	5,5-diCH ₃	OCH ₃	6-OCH ₃	7.26
B-3^a	5,5-diCH ₃	OCH ₃	7-OCH ₃	6.52
B-4	5,5-diCH ₃	OCH ₃	6-Cl	7.26
B-5	5,5-diCH ₃	OCH ₃	7-SCH ₃	6.47
B-6	5,5-diCH ₃	OCH ₃	8-CH ₂ CH ₃	6.65
B-7	5,5-diCH ₃	OCH ₃	8-CH(CH ₃) ₂	6.88
B-8	5,5-diCH ₃	OCH ₃	7-F-8-CH ₃	6.20
B-9^a	5,5-diCH ₃	OCH ₃	7-Cl-8-CH ₃	6.54
B-10	5,5-diCH ₃	OCH ₃	7-Br-8-CH ₃	7.39
B-11	5,5-diCH ₃	OCH ₃	8-CH ₃	6.53
B-12	5-CH ₃	OCH ₃	8-CH ₃	7.00
B-13^a	4,4-diCH ₃	OCH ₃	8-CH ₃	7.13
B-14^a	H	OCH ₃	8-CH ₃	8.05
B-15	H	OCH ₂ H ₃	8-CH ₃	7.60
B-16	H	O- <i>n</i> -Pr	8-CH ₃	6.77
B-17	H	OCH ₂ Ph	8-CH ₃	6.67
B-18	H	OCH ₃	H	7.21
B-19	H	OCH ₃	6-CH ₃	7.52
B-20	H	OCH ₃	7-CH ₃	7.62
B-21	H	OCH ₃	6-OCH ₃	7.12
B-22	H	OCH ₃	7-OCH ₃	7.54
B-23	H	OCH ₃	6-Cl	7.14
B-24	H	OCH ₃	7-Cl	7.96

B-25^a	H	OCH ₃	7-SCH ₃	8.05
B-26^a	H	OCH ₃	8-CH ₂ CH ₃	7.77
B-27^a	H	OCH ₃	8-CH(CH ₃) ₂	7.46
B-28	H	OCH ₃	5,8-diCH ₃	7.41
B-29	H	OCH ₃	6,8-diCH ₃	7.55
B-30	H	OCH ₃	7,8-diCH ₃	7.80
B-31^a	H	OCH ₃	7-F-8-CH ₃	7.96
B-32	H	OCH ₃	7-Cl-8-CH ₃	7.85
B-33^a	H	OCH ₃	7-Br-8-CH ₃	7.68
B-34	H	CH ₃	8-CH ₃	8.15
B-35	5-CH ₃	CH ₃	8-CH ₃	8.15
B-36^a	H	CF ₃	8-CH ₃	7.70
B-37	H	CN	8-CH ₃	8.15
B-38	5-CH ₃	CN	8-CH ₃	8.10
Mesotrione^a				7.89

^a Randomly selected test set compounds.

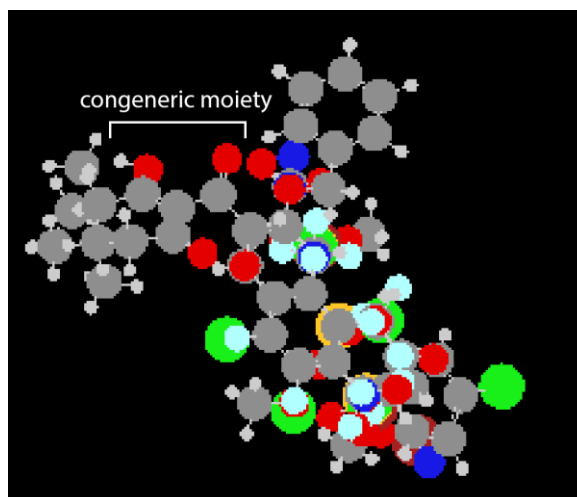


Figure 2. Superposition of the 97 triketone chemical images illustrating the congeneric scaffold and the variable moieties within the data set.

To validate the MIA-QSAR outcomes, as well as to understand the ligand-enzyme interactions responsible for the experimental and predicted bioactivities, we performed molecular docking simulations using the crystal structure of *A*tHPPD (EC 1.13.11.27) complexed to the inhibitor 1,5-dimethyl-6-(2-oxidanyl-6-oxidanylidene-cyclohexen-1-

yl)carbonyl-3-(3-trimethylsilylprop-2-ynyl)quinazoline-2,4-dione (NJK). This complex was obtained from the Protein Data Bank (PDB code: 6M6D, resolution: 1.80 Å) and Co^{II} rather than Fe^{II} was the ion found in the binding site (Qu et al. 2021), which differs from the protein used elsewhere (Wang et al. 2015; 2016). Hydrogen atoms were added, water molecules were removed, and structure refinement was performed using the steepest descents and the conjugate gradients algorithms, until the final root-mean-square energy gradient was less than 0.10 kcal mol⁻¹. In order to remove close contacts after adding the hydrogens, heavy atoms were initially constrained, while hydrogens were relaxed; then, the backbone was constrained and the side chains were relaxed; finally, all the constraints were removed and the whole molecule was relaxed. The ligands were optimized using the AMBER force field and atomic charges were calculated at the AM1-BCC level. The protein and ligand structures were prepared using the Discovery Studio for further docking simulations using the Molegro Virtual Docker software (Thomsen and Christensen 2006). The algorithm used for cavity prediction and recognition of potential binding modes, the poses, is derived from the piecewise linear potential, whose parameters are fit to protein-ligand structures and binding data scoring functions and further extended in GEMDOCK program (Generic Evolutionary Method for molecular DOCK), which includes a hydrogen bonding term and new charge schemes (Thomsen and Christensen 2006). Accordingly, a grid of 0.30 Å resolution and 10 Å radius from the inhibitor as the binding site were applied. MolDock SE (Simplex Evolution) was used as the search algorithm, and the maximum number of poses to generate was 100.

Results and Discussion

QSAR has demonstrated to be a key tool for the discovery and development of drugs and agrochemicals (Todeschini et al. 2020). However, the QSAR models should be strongly validated to attest for their predictive capability and reliability. Accordingly, the three MIA-QSAR models based on r_{vdW} , ϵ , and r_{vdW}/ϵ -encoding MIA descriptors were evaluated with regard to their modeling performance (acceptable cut off: $r^2 \geq 0.6$), chance correlation (acceptable cut off: ${}^c r_p^2 \geq 0.5$), predictive power (acceptable cut off: q^2 and $r_{pred}^2 \geq 0.5$), and accuracy (acceptable cut off: avg. $r_m^2 \geq 0.5$ and $CCC \approx 0.85$). Table 2 shows that the three models exhibited similar performance and, therefore, both atomic radius (related to steric effects) and electronegativity (related to electrostatic interactions in the binding pocket) are able to explain the bioactivity data. Moreover, the three models demonstrated high prediction

performance and reliability (Figure 3). Therefore, the PLS regression parameters of these models can be used to estimate the pK_i values of hitherto unknown congeners.

Table 2. Statistical parameters for the MIA-QSAR models.

Parameter	r_{vdW}	ε	r_{vdW}/ε	Average	Std. Dev.
PLS comp.	3	3	3	3	0.00
RMSEC	0.40	0.38	0.41	0.40	0.02
r^2	0.71	0.74	0.70	0.72	0.02
RMSEy-rand	0.59	0.59	0.61	0.60	0.01
$r^2_{y\text{-rand}}$	0.37	0.38	0.34	0.36	0.02
$c_r^2_p$	0.50	0.52	0.50	0.51	0.01
RMSECV	0.54	0.54	0.54	0.54	0.00
q^2	0.51	0.51	0.51	0.51	0.00
RMSEP	0.43	0.44	0.42	0.43	0.01
r^2_{pred}	0.71	0.69	0.73	0.71	0.02
Avg. r^2_m	0.61	0.58	0.64	0.61	0.03
Δr^2_m	0.07	0.08	0.04	0.06	0.02
CCC	0.83	0.82	0.84	0.83	0.01

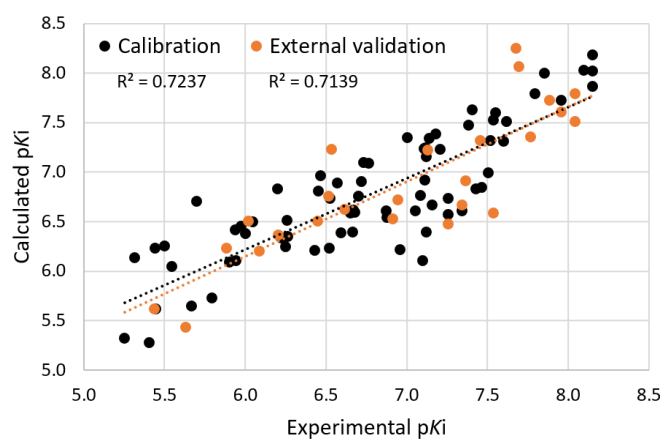


Figure 3. Correlation plot of experimental vs. predicted (mean values) pK_i for the calibration and external validation sets of substituted triketones.

To guide the proposition of new compounds, it is important to understand the role of substituents to the bioactivity data. In this sense, the PLS regression coefficients (**b**) indicate the negative or positive influence of each MIA descriptor to the herbicidal activity, while variable importance in projection (VIP) scores describe the strength of this influence. These parameters for the whole series of molecules were plotted to give the contour maps of Figure 4. In the VIP-plot, the substituent contribution to pK_i increases as the atom colors range from blue to red; in the **b**-plot, atomic color tending to blue decreases the bioactivity, while tending to red contributes to increase the pK_i values. Since the general profiles of the VIP and **b**-plots for the r_{vdW} , ϵ , and r_{vdW}/ϵ -based contour maps are the same, a unique analysis is provided herein. The VIP-plot suggests that CH_3 in R^1 , C and N atoms of the quinoline ring, and OCH_3 in R^2 of scaffold **B** contribute importantly to the bioactivity (yellow to red regions in the plot). Secondary effects are observed for halogens, especially F in R^3 of scaffold **A1** and 7-Cl in R^3 of scaffold **B**. In turn, the **b**-plot indicates that CH_3 in R^1 is detrimental to pK_i , while 7-Cl in R^3 and CN in R^2 of scaffold **B** contribute to increase the herbicidal activity; 5-Cl and fluorines in R^3 are also advantageous. Based on these findings and combining the most promising substituents (those leading to higher pK_i values in the library compounds) in the scaffold **B**, we propose 28 chemical candidates as shown in Table 3.

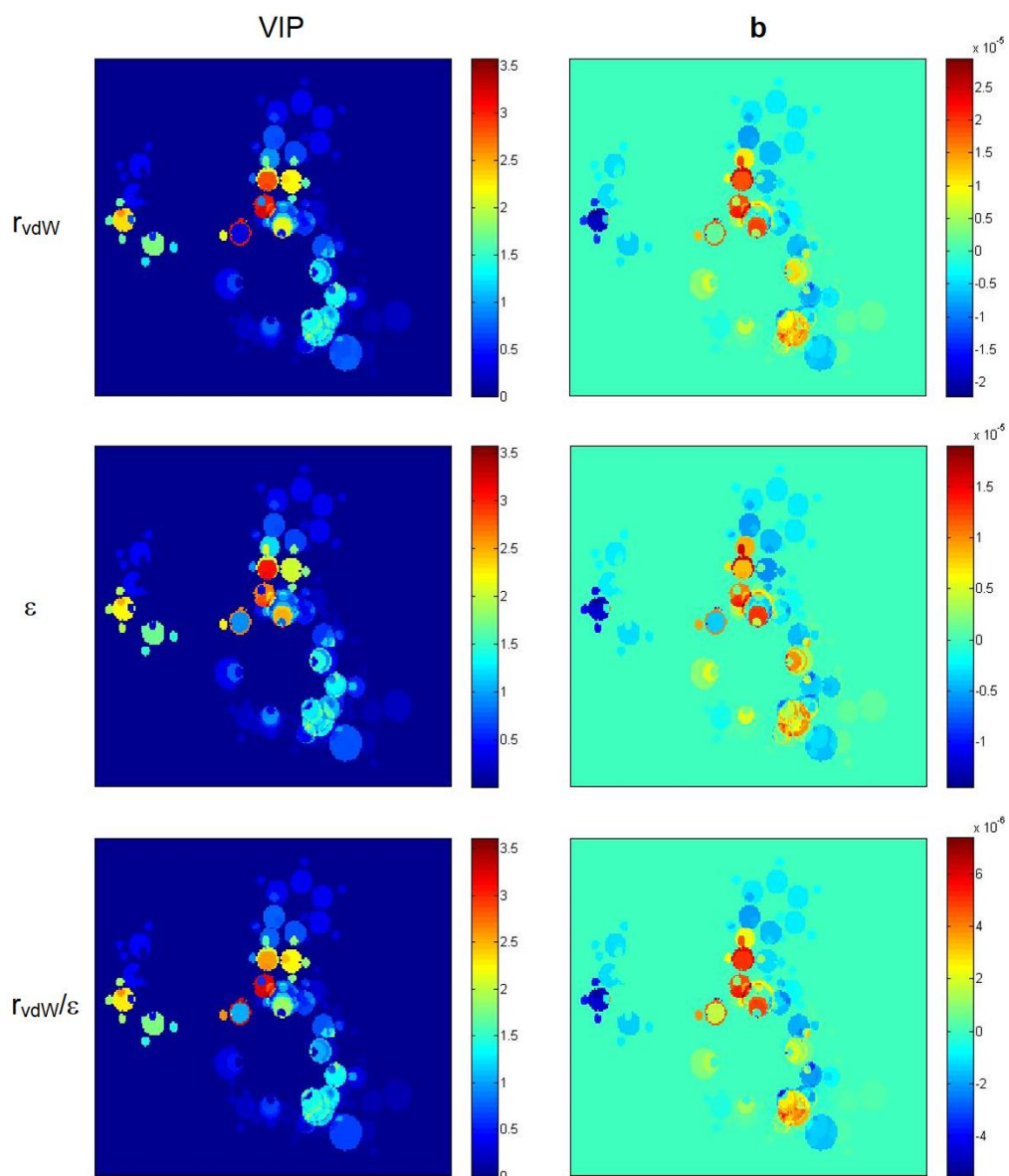


Figure 4. VIP and **b**-plots for the MIA-QSAR models based on r_{vdW} , ϵ , and r_{vdW}/ϵ - encoding descriptors. Refer to Figure 2 for comparison of the chemical structures with the MIA contour maps.

Table 3. Proposed triketones with scaffold **B** and the respective pK_i values predicted using the MIA-QSAR models.

Compound	R ¹	R ²	R ³	r_{vdw}	ϵ	r_{vdw}/ϵ	Avg.	Std. Dev.
P1	H	CH ₃	5,6,7,8-F	8.09	8.37	8.03	8.16	0.18

P2	H	CH ₃	5-Cl	7.92	7.91	7.97	7.93	0.03
P3	H	CH ₃	8-Cl	8.02	8.12	7.99	8.05	0.07
P4	H	CH ₃	7-Br	8.51	8.41	8.53	8.48	0.06
P5	H	CH ₃	7-NO ₂	8.14	8.22	8.14	8.17	0.05
P6	H	CH ₃	7-Cl	8.24	8.17	8.29	8.23	0.06
P7	H	CH ₃	7-SCH ₃	8.26	8.10	8.40	8.25	0.15
P8	H	CH ₃	7-CF ₃	8.10	8.06	8.15	8.10	0.04
P9	H	CH ₃	5,7,8-Cl	8.74	8.87	8.58	8.73	0.14
P10	H	CN	5,6,7,8-F	8.24	8.64	8.10	8.33	0.28
P11	H	CN	5-Cl	8.07	8.18	8.03	8.09	0.08
P12	H	CN	8-Cl	8.18	8.39	8.05	8.21	0.17
P13	H	CN	7-Br	8.66	8.68	8.59	8.64	0.04
P14	H	CN	7-NO ₂	8.29	8.49	8.20	8.33	0.15
P15	H	CN	7-Cl	8.40	8.44	8.35	8.40	0.04
P16	H	CN	7-SCH ₃	8.41	8.36	8.46	8.41	0.05
P17	H	CN	7-CF ₃	8.25	8.33	8.21	8.26	0.06
P18	H	CN	5,7,8-Cl	8.89	9.14	8.64	8.89	0.25
P19	H	H	5,6,7,8-F	7.48	7.84	7.40	7.58	0.23
P20	H	H	5-Cl	7.31	7.38	7.34	7.34	0.03
P21	H	H	8-Cl	7.42	7.59	7.36	7.46	0.12
P22	H	H	7-Br	7.90	7.88	7.90	7.89	0.01
P23	H	H	7-NO ₂	7.53	7.69	7.51	7.58	0.10
P24	H	H	7-Cl	7.64	7.63	7.66	7.64	0.01
P25	H	H	7-SCH ₃	7.65	7.56	7.77	7.66	0.10
P26	H	H	7-CF ₃	7.49	7.53	7.52	7.51	0.02
P27	H	H	5,7,8-Cl	8.13	8.33	7.95	8.14	0.19
P28	H	H	8-CH ₃	7.40	7.49	7.40	7.43	0.05

The first 14 best molecules of Figure 5 are proposed compounds, since they possess estimated pK_i larger than the library compounds. Compound **P18** is particularly recommended for further synthesis (Scheme 1) because of its high mean pK_i value of 8.89, better than that of mesotrione; the candidate molecule **P18** contains $R^1 = H$, $R^2 = CN$, $R^3 = 5,7,8\text{-triCl}$, and the backbone **B**, in agreement to the guidance provided by the MIA-plots. However, since these

compounds are particularly useful for post-emergence application, their efficacy should be also dependent on the leaf uptake. Such absorption depends on the lipophilicity, described in terms of the octanol-water partition coefficient, $\log P$. Although high lipophilicity may provide ecological issues, such as bioaccumulation, it may be advantageous as lipophilic compounds have lower mobility in soils rich in organic matter and, therefore, are less prone to leach out and contaminate the groundwater (Martins et al. 2022). According to Figure 5, compound **P18**, as well as most of the library and proposed compounds, are much more lipophilic than mesotrione.

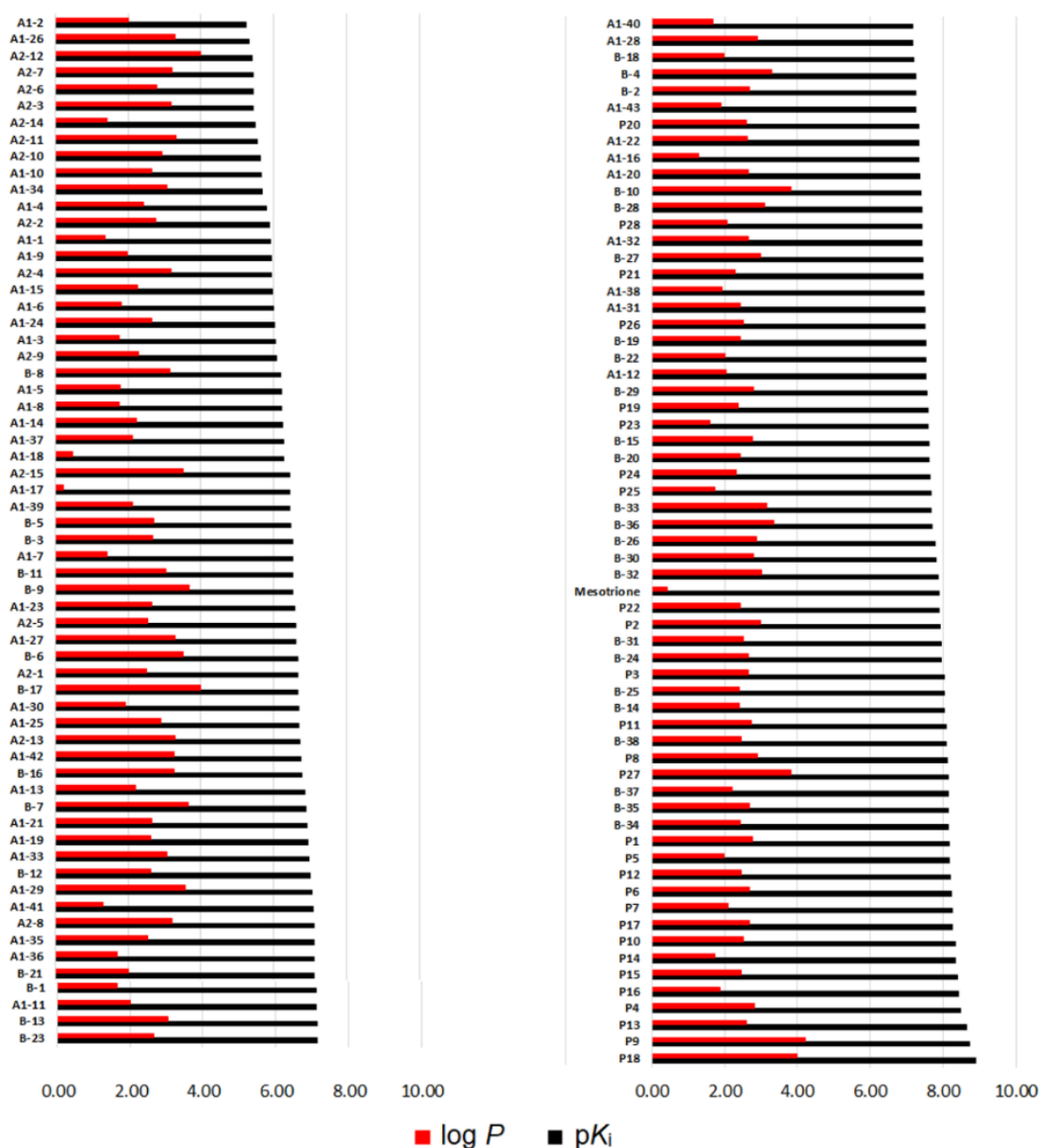
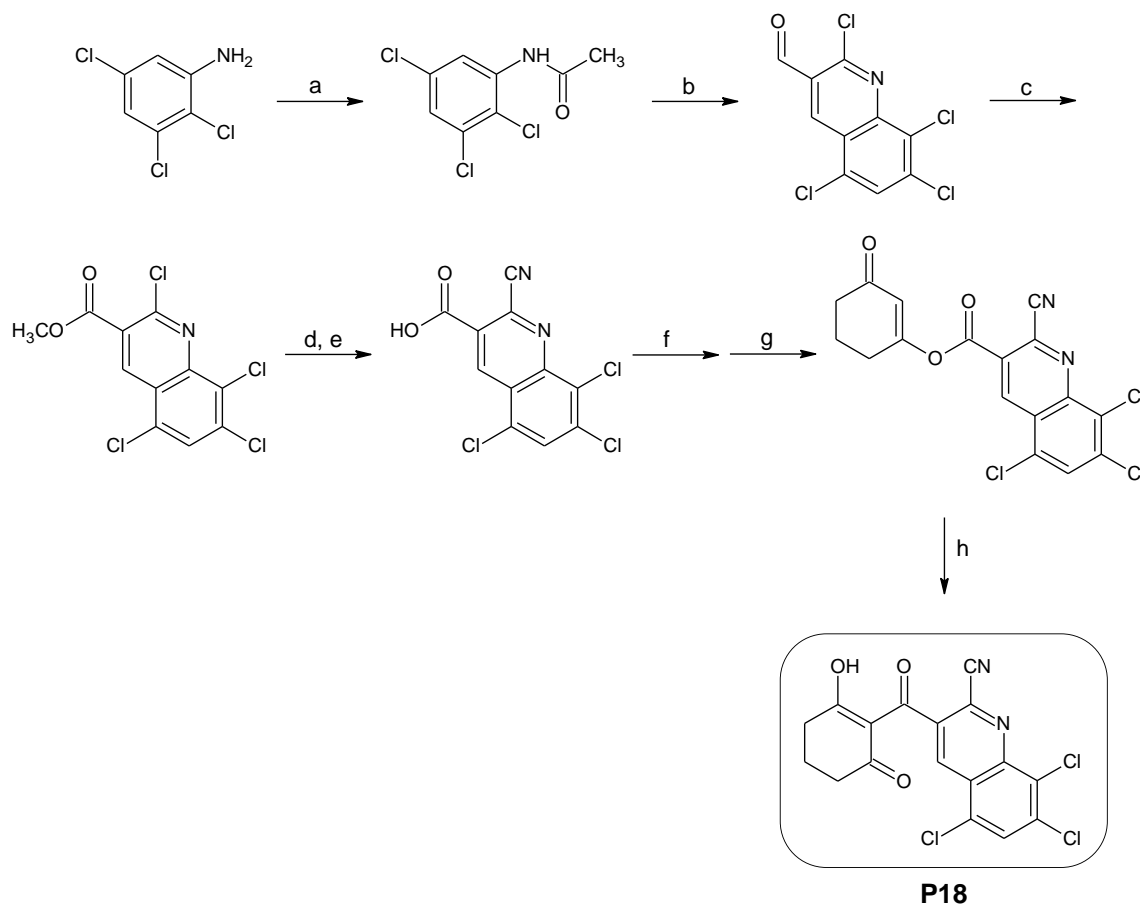


Figure 5. Calculated/predicted $\log P$ and pK_i for the library and proposed triketone derivatives.



Scheme 1. Possible synthetic route to obtain **P18**, similar to the procedure described by Wang et al. (2015). Reagents and conditions: (a) Ac_2O , Et_3N , CH_2Cl_2 , 0°C –room temperature; (b) DMF, POCl_3 , 0 – 80°C ; (c) K_2CO_3 , CH_3OH , I_2 , reflux; (d) CuCN , DMF, reflux; (e) H_2O , citric acid; (f) COCl_2 , -5°C ; (g) 1,3-cyclohexanedione, CH_2Cl_2 , 0°C ; (h) acetone cyanohydrin, Et_3N , CH_2Cl_2 , room temperature.

To further explore the interactions responsible for the herbicidal activities, molecular docking studies were performed on the representative compounds **A1-2** (lowest activity value), **B-1** (intermediate activity value), **B-34** and **B-37** (highest activity values), **P-18** (proposed), and mesotrione. The docking protocol was validated for *At*HPPD by removing the inhibitor (**NJX**) from the complex, re-docking, and calculating the root mean square deviation (RMSD) for two conformations: the original from the Protein Data Bank (**NJX-1**) and the energy minimized conformation (**NJX-2**). Both poses yielded RMSDs lower than 1.5 to 2.0 Å, that is, 0.41 Å for **NJX-1** and 0.49 Å for **NJX-2**, and, therefore, the procedure can be considered successful (Pagadala et al. 2017). The same parameters were used in the ligand-receptor docking process. The overall orientation of these five molecules within the active site

pocket of *At*HPPD is shown in Figure 6, which depicts a similar binding mode to that of the co-crystallized ligand (**NJX**). Table 4 shows the interaction energy values obtained from the molecular docking, whereas Figure 7 illustrates the interactions between the compounds and *At*HPPD responsible for the docking scores. Both benzoyl or quinoline moieties establish π - π stacking interactions with Phe424 and Phe381, but it is possible that compounds with scaffold **B** have better *At*HPPD inhibitory activity than scaffold **A** due to a greater stabilization caused by the interaction of the fused rings of quinoline with the two amino acids. Also, the complex formed with **B-1**, **B-34**, and **B-37** is more stable than with **A1-2** because of more effective interactions with the cofactor (Table 4). Furthermore, **A1-2** does not exhibit any hydrogen bond with other amino acids. In turn, **B-34** and **B-37** engage in hydrogen bonds with Gln307, His308, and His226, and show better docking score values than the other compounds in Table 4 (except **P-18**). The proposed compound **P-18** establishes hydrogen bonds with Gln307 and His308, and halogen interactions with Gln379 and Gly420. From Figure 8, in which the good correlation between docking scores and predicted pK_i for the proposals is depicted ($R^2 = 0.70$), it appears that **P4**, **P7**, **P9**, and **P12-P18** are more stable in *At*HPPD than mesotrione, **NJX**, and **B-34**. However, although **P18** possesses a lower (more stable) interaction energy with cobalt, **NJX** presents a higher intermolecular interaction and forms a more stable complex with the water⁶⁴⁴ molecule (Table 4); the distances from the 1,3-diketone moiety to the cobalt and water⁶⁴⁴ are within the range of 1.91-3.03 Å (Figure 7).

Table 4. Binding energy values (kcal mol⁻¹) of the docked ligands into the *At*HPPD binding pocket.

Compound	MolDock Score	Intermolecular	Cobalt	Water ⁶⁴⁴	H-bond
A1-2	-88.12	-112.06	-0.18	-4.85	0
B-1	-94.76	-117.12	-2.10	-4.73	-2.94
B-37	-105.02	-127.70	-2.40	-4.73	-5.01
B-34	-107.38	-128.31	-2.35	-4.72	-3.18
P-18	-121.04	-119.30	-3.17	-4.62	-3.00
Mesotrione	-101.91	-124.25	-1.74	-4.65	-0.29
NJX	-116.68	-145.99	-1.40	-7.38	0

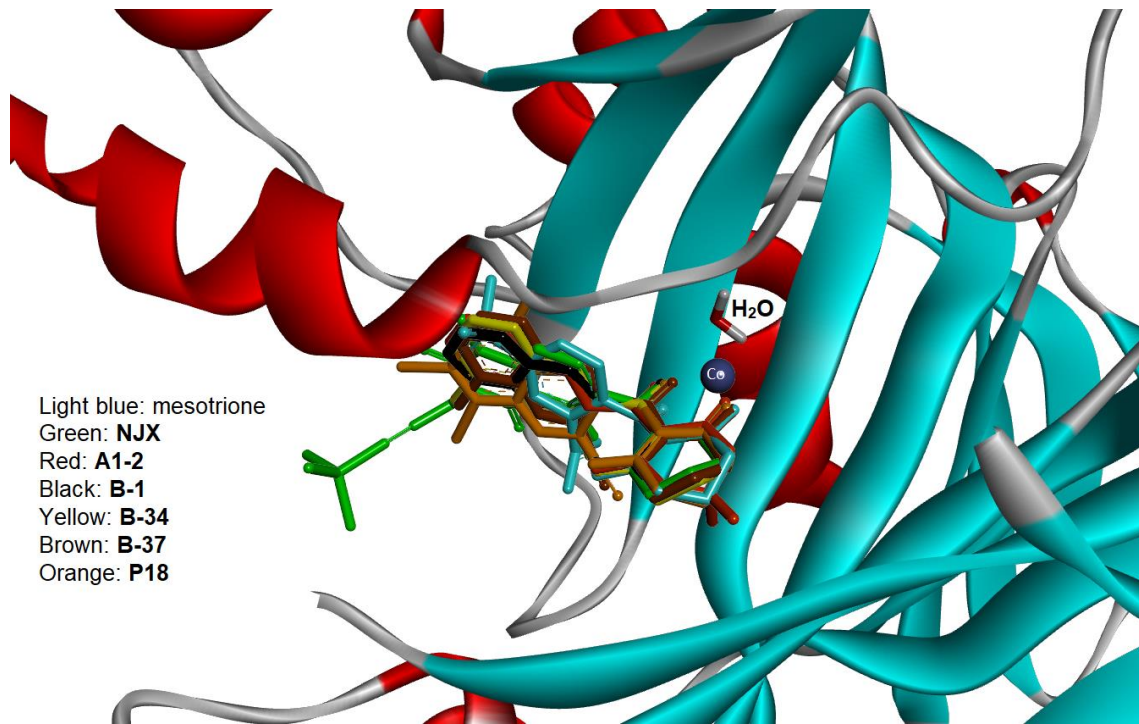


Figure 6. Superimposed view of *At*HPPD and mesotrione, NJX, A1-2, B-1, B-34, B-37 and P18.

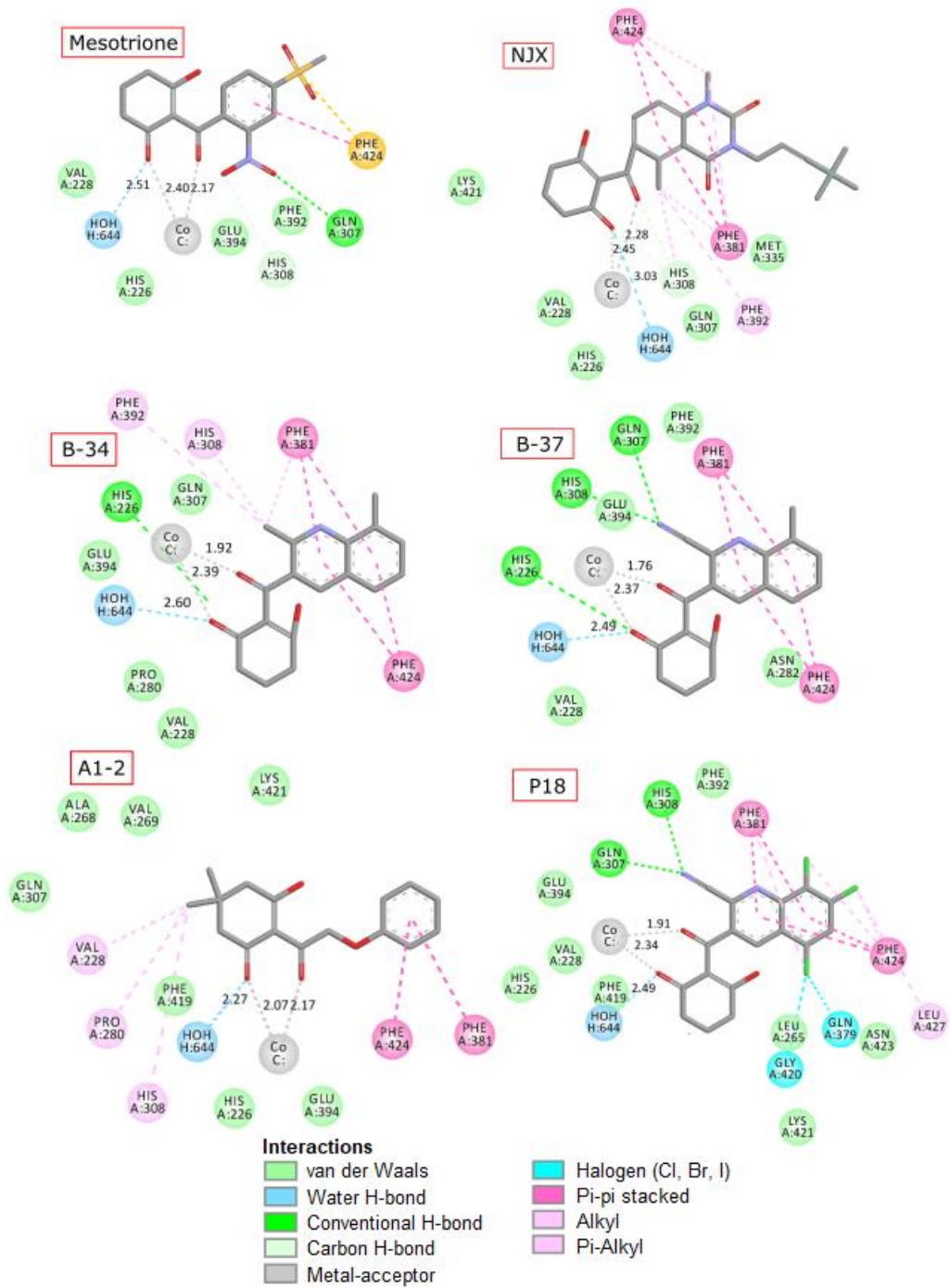


Figure 7. 2D theoretical insight of the interactions between selected compounds and the *AtHPPD* enzyme.

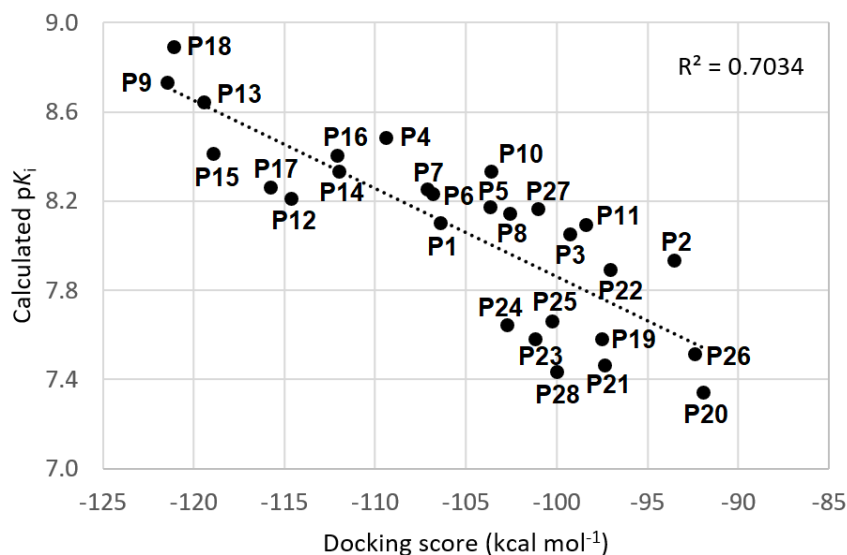


Figure 8. Correlation plot of docking scores (kcal mol⁻¹) *versus* calculated pK_i for the proposed compounds **P1-P28**.

Acknowledgments

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, funding code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant number 306830/2021-3), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for financial support of this research.

Author Contribution Statement

L. R. Capucho: investigation; E. F. F. da Cunha: investigation, review & editing; M. P. Freitas: original draft, supervision.

Conflicts of interest

There are no conflicts to declare.

References

Barigye SJ, Duarte MH, Nunes CA, Freitas MP (2016) MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. *RSC Adv* 6:49604-49612.

- Beaudegnies R, Edmunds AJF, Fraser TEM, Hall RG, Hawkes TR, Mitchell G, Schaezter J, Wendeborn S, Wibley J (2009) Herbicidal 4-hydroxyphenylpyruvate dioxygenase inhibitors – A review of the triketone chemistry story from a Syngenta perspective. *Bioorg Med Chem* 17:4134–4152.
- 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.
- Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q^2 parameter for QSAR validation. *J Chem Inf Model* 49:1669-1678.
- Dennington RD, Keith TA, Millam MJ (2008) GaussView 5.0. Wallingford, CT.
- Freitas MR, Barigye SJ, Freitas MP (2015) Coloured chemical image-based models for the prediction of soil sorption of herbicides. *RSC Adv* 5:7547-7553.
- Geladi P, Kowalski BR (1986) Partial least-squares regression: A tutorial. *Anal Chim Acta* 185:1-17.
- Kaundun SS, Hutchings S-J, Dale RP, Howell A, Morris JA, Kramer VC, Shivrain VK, Mcindoe E (2017) Mechanism of resistance to mesotrione in an *Amaranthus tuberculatus* population from Nebraska, USA. *Plos One* 12:e0180095.
- Martins, FA, Daré JK, Freitas MP (2022) Computer-assisted proposition of promising aryloxyacetic acid derivatives as HPPD inhibitors. *J Agric Food Chem* 70:8986-8993.
- 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.
- Moran GR (2005) 4-Hydroxyphenylpyruvate dioxygenase. *Arch Biochem Biophys* 433:117-128.
- Nunes CA, Freitas MP, Pinheiro ACM, Bastos SC (2012) Chemoface: a novel free user-friendly interface for chemometrics. *J Braz Chem Soc* 23:2003-2010.
- Oliveira MC, Gaines TA, Jhala AJ, Knezevic SZ (2018) Inheritance of mesotrione resistance in an *Amaranthus tuberculatus* (var. *rudis*) population from Nebraska, USA. *Front Plant Sci* 9:60.
- Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. *Biophys Rev* 9:91-102.
- Qu RY, Nan JX, Yan YC, Chen Q, Ndikuryayo F, Wei XF, Yang WC, Lin HY, Yang GF (2021) *J Agric Food Chem* 69:459-473.

- Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) Some case studies on application of “ r_m^2 ” metrics for judging quality of quantitative structure–activity relationship predictions: Emphasis on scaling of response data. *J Comput Chem* 34:1071-1082.
- Thomsen R, Christensen MH (2006) MolDock: A new technique for high-accuracy molecular docking. *J Med Chem* 49:3315-3321.
- Todeschini R, Consonni V, Ballabio D, Grisoni F (2020) Chemometrics for QSAR Modeling. In: Brown SD, Tauler R, Walczak B (eds.) *Comprehensive Chemometrics*. Elsevier, Amsterdam, pp 599-634.
- Wang DW, Lin H-Y, Cao R-J, Chen T, Wu F-X, Hao G-F, Chen Q, Yang W-C, Yang G-F (2015) Synthesis and herbicidal activity of triketone–quinoline hybrids as novel 4-hydroxyphenylpyruvate dioxygenase inhibitors. *J Agric Food Chem* 63:5587-5596.
- Wang DW, Lin H-Y, He B, Wu F-X, Chen T, Chen Q, Yang W-C, Yang G-F (2016) An efficient one-pot synthesis of 2-(aryloxyacetyl)cyclohexane-1,3-diones as herbicidal 4-hydroxyphenylpyruvate dioxygenase inhibitors. *J Agric Food Chem* 63:5587-5596.