



## Predição de Propriedades Toxicológicas de Ingredientes Farmacêuticos Ativos através de Metodologia QSAR 2D

Almir P. D. S. Filho<sup>1\*</sup> (IC); Eduardo B. de Melo<sup>2</sup> (PQ) João P. A. Martins<sup>1\*</sup> (PQ).

<sup>1</sup> Departamento de Química/ICEx, Universidade Federal de Minas Gerais (UFMG), Departamento de Química/ICEx, Belo Horizonte, MG, Brasil.

<sup>2</sup> Curso de Farmácia, Universidade Estadual do Oeste do Paraná (UNIOESTE)

[pachecoalmir@hotmail.com](mailto:pachecoalmir@hotmail.com)

[joaopauloam@ufmg.br](mailto:joaopauloam@ufmg.br)

**Palavras-Chave:** Toxicidade, Contaminantes, Descritores.

### Introdução

A presença de resíduos farmacêuticos no ambiente aquático tem se tornado uma preocupação crescente nas últimas décadas, em razão de seus potenciais efeitos tóxicos sobre organismos não-alvo, como algas, invertebrados e peixes. Muitos fármacos não são totalmente degradados pelos sistemas convencionais de tratamento de esgoto, sendo frequentemente detectados em corpos hídricos em concentrações que variam de ng/L a mg/L — níveis suficientes para provocar impactos ecotoxicológicos relevantes. Os Ingredientes Farmacêuticos Ativos (IFAs), devido à sua persistência e atividade biológica, vêm sendo reconhecidos desde a década de 1970 como Contaminantes Emergentes (CECs) de grande interesse ambiental.

A avaliação experimental da toxicidade desses compostos, embora essencial, envolve elevados custos, grande consumo de tempo e o uso extensivo de organismos-teste. Nesse contexto, a metodologia QSAR (Quantitative Structure–Activity Relationship) surge como uma alternativa eficiente, ao estabelecer correlações entre características físico-químicas, eletrônicas e topológicas das moléculas — representadas por descritores moleculares calculados a partir da estrutura química — e atividades biológicas de interesse (Martins et. al. 2009), como a toxicidade aguda (pLC<sub>50</sub>, 96h), que indica a concentração necessária para causar mortalidade em 50% da população experimental. Essa abordagem permite prever propriedades de substâncias ainda não testadas, preenchendo lacunas de dados ausentes através da relação estabelecida.

O presente estudo baseia-se no artigo “*Hazard of pharmaceuticals for aquatic environment: Prioritization by structural approaches and prediction of ecotoxicity*”, publicado por Sangion e Gramatica (2016), que propôs uma estratégia de priorização de IFAs por meio de modelos QSAR QSAR 2D). Inspirados nessa abordagem, avaliamos propriedades toxicológicas de fármacos por meio de modelos estatísticos validados, capazes de prever a toxicidade aguda em diferentes organismos aquáticos. Assim, desenvolvemos um modelo QSAR 2D utilizando descritores moleculares selecionados via algoritmo genético, avaliando sua robustez e aplicabilidade em comparação com o estudo original, com ênfase na espécie de peixe *Pimephales promelas*. Dessa forma, buscou-se contribuir para a priorização de substâncias potencialmente perigosas ao ambiente aquático.

### Material e Métodos

O conjunto de dados utilizado neste trabalho foi obtido a partir do artigo de referência, consistindo em 62 moléculas descritas em notação SMILES (Simplified Molecular Input Line



Entry System), acompanhadas dos respectivos valores de toxicidade aguda expressos em  $pLC_{50}$  ( $mmol L^{-1}$ ).

Inicialmente, as estruturas moleculares em formato SMILES foram processadas pelo programa em Python “runxtb.py”, que gera arquivos no formato xyz utilizando o pacote OpenBabel e realiza uma otimização preliminar de geometria utilizando o software xtb com o método GFN2-xTB. Em seguida, essas geometrias otimizadas foram submetidas a uma segunda etapa de refinamento, mais robusta, com o programa “runpsi4.py”, que emprega a biblioteca Psi4. Nessa etapa, foram realizados cálculos de química quântica com base na Teoria do Funcional da Densidade (DFT), utilizando o funcional B3LYP e a base cc-pVTZ, para obtenção de energias e propriedades moleculares adicionais.

Os arquivos resultantes em formato xyz foram convertidos para mol2 por meio do OpenBabel, possibilitando o cálculo de descritores moleculares. Para esse fim, empregou-se o pacote Mordred, que faz uso da biblioteca RDKit, e gera um grande número de descritores. A matriz de descritores obtida foi posteriormente padronizada com a função StandardScaler da biblioteca scikit-learn (sklearn.preprocessing), garantindo comparabilidade entre variáveis.

A matriz de descritores foi reduzida eliminando-se descritores redundantes (correlação entre si  $\geq 0,9$ ) e uma seleção de variáveis por meio de um algoritmo genético (GA) foi realizada a fim de identificar aqueles descritores com maior relevância para a atividade biológica. Esse procedimento assegurou a construção de um conjunto otimizado de descritores para o modelo, seguido de validação cruzada, y-randomization e leave-N-out para testar sua robustez, além de uma validação externa, utilizando-se um conjunto teste contendo 25 % das amostras, selecionadas pelo algoritmo de Kennard-Stone. Todos os programas e algoritmos mencionados foram construídos em nosso grupo de pesquisa.

Finalmente, o modelo QSAR 2D foi ajustado por meio de regressão por mínimos quadrados parciais (PLS), técnica adequada para lidar com colinearidade entre descritores. Foram utilizadas quatro variáveis latentes ( $nLV = 4$ ), permitindo o ajuste eficiente aos dados de treinamento. O modelo resultante foi então aplicado à predição da toxicidade dos compostos de teste, gerando gráficos de valores experimentais versus preditos. Esses resultados confirmam a aplicabilidade do modelo para a priorização de Ingredientes Farmacêuticos Ativos (IFAs) potencialmente perigosos ao ambiente aquático.

## Resultados e Discussão

O modelo QSAR 2D desenvolvido apresentou desempenho estatístico robusto, tanto na etapa de calibração quanto nas etapas de validação (ROY et. al., 2016). Para o conjunto de treinamento, os resultados indicaram excelente ajuste, com coeficiente de determinação ( $R^2$ ) = 0,959, erro quadrático médio da calibração (RMSEC) = 0,342 e correlação no conjunto de calibração ( $r_{cal}$ ) = 0,979. Além disso, os índices  $avgR^2_{cal} = 0,94$  e  $\Delta R^2_{cal} = 0,037$  reforçam a confiabilidade do modelo.

Na validação cruzada, foi observado  $Q^2 = 0,925$ , com  $RMSECV = 0,462$  e  $r_{cv} = 0,962$ , evidenciando a capacidade preditiva do modelo e sua generalização. A estatística F ( $F = 232,42$ ) indica que o modelo é estatisticamente significativo ( $p < 0,05$ ).

Adicionalmente, a validação externa corroborou a robustez do modelo, apresentando valores consistentes sob diferentes divisões do conjunto de dados ( $Q^2F1 = 0,77$ ;  $Q^2F2 = 0,68$ ;  $Q^2F3 = 0,81$ ;  $Q^2F195 = 0,85$ ;  $Q^2F295 = 0,80$ ;  $Q^2F395 = 0,89$ ). O erro médio de predição foi de (RMSEP = 0,729; MAE = 0,533).

Quando comparado ao modelo do estudo de referência, que obteve  $R^2 = 0,80$  e  $Q^2 = 0,75$ , o modelo aqui proposto demonstrou desempenho superior. Essa diferença pode ser atribuída, em parte, ao uso de cálculos quânticos (DFT/B3LYP cc-pVTZ), capazes de capturar propriedades eletrônicas detalhadas — como distribuição de carga, orbitais moleculares e interações intramoleculares —, aspectos não contemplados no modelo original, que se baseou exclusivamente em descritores obtidos pelo software PaDEL-Descriptor.

Tabela 1 – Tabela comparativa dos parâmetros estatísticos dos modelos

|                          | $R^2$ | $Q^2$ | $Q^2_{\text{ext}} F_n$ | $\text{RMSE}_{\text{ext}}$ | $\text{MAE}_{\text{ext}}$ |
|--------------------------|-------|-------|------------------------|----------------------------|---------------------------|
| Modelo artigo referência | 0,80  | 0,73  | 0,70-0,78              | 0,84                       | 0,77                      |
| Modelo desenvolvido      | 0,96  | 0,92  | 0,68-0,81              | 0,73                       | 0,53                      |

A Figura 1 mostra a comparação entre os valores preditos e valores experimentais. O conjunto de teste consta com 15 moléculas, representando 25% do total de moléculas estudadas.

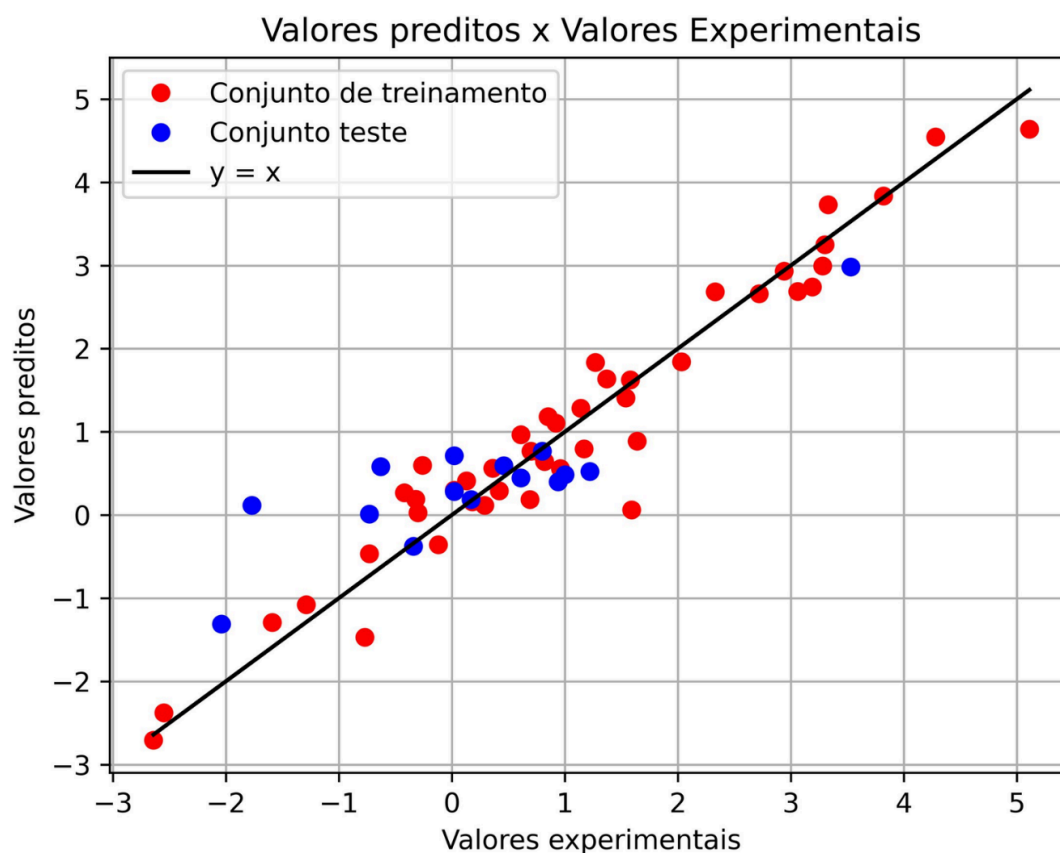


Figura 1 – Valores Preditos x Valores Experimentais

Entre as moléculas com maior toxicidade destacam-se: Cis-Permethrin, Hexachlorophene, Triclocarban, Hexachlorocyclohexane, Sertraline e (+)-Fluoxetine. O algoritmo de seleção identificou 18 descritores relevantes, apresentados na Tabela 2. Entretanto, nesta etapa do estudo, ainda não foi possível estabelecer de forma conclusiva a



relação individual nem o grau de influência de cada descritor sobre as toxicidades biológicas das moléculas analisadas. Para tal, seria necessária uma investigação mais aprofundada, com aplicação de ferramentas adicionais de análise multivariada e interpretação química.

Tabela 2 – Descritores mais relevantes

| Categoria / Grupo                      | Descritores  |
|--|--|
| Momentos multipolares                  | Dipole Polarizability XY, Quadrupole XY, Quadrupole YY |
| Autocorrelações topológicas            | ATSC5m, ATSC5v, AATSC4c, GATS2d, GATS2m                |
| Estrutura e conectividade              | C3SP3, ZMIC1, JGII                                     |
| Grupos funcionais e átomos             | nBase, NdsCH, SsssN                                    |
| Propriedades eletrônicas e superfícies | PEOE_VSA12, VSA_EState9, AXp-1dv                       |
| Anéis aromáticos                       | nFHRing  |

## Conclusões

Com base na análise realizada no estudo QSAR 2D, foi possível desenvolver um modelo robusto e altamente preditivo para estimar a toxicidade aquática de fármacos a partir de suas estruturas moleculares. O modelo apresentou excelentes métricas estatísticas, como um coeficiente de determinação ( $R^2$ ) de 0,959 e um coeficiente de validação cruzada ( $Q^2$ ) de 0,925, superando significativamente os valores obtidos no artigo de referência ( $R^2 = 0.8$  e  $Q^2 = 0.76$ ). Isso demonstra a capacidade do modelo em explicar a variabilidade dos dados e generalizar para novas moléculas.

A metodologia empregada, que incluiu a seleção de uma gama vasta de descritores moleculares diversificados (como momentos multipolares, autocorrelações topológicas e propriedades eletrônicas), otimização geométrica e cálculo de descritores via DFT, validação interna e externa, garantiu maior confiabilidade e aplicabilidade do modelo. Além disso, a comparação entre valores preditos e experimentais mostrou uma forte correlação, reforçando a precisão das previsões.

Os resultados obtidos destacam a importância de ferramentas *in silico*, como o QSAR 2D, para priorizar fármacos potencialmente perigosos ao ambiente aquático, reduzindo a necessidade de testes em animais e otimizando recursos. O modelo desenvolvido neste trabalho mostrou-se válido, eficiente e aplicável, podendo ser utilizado para prever a toxicidade de moléculas em organismos aquáticos, contribuindo para a avaliação de riscos ambientais e a tomada de decisões regulatórias.

Portanto, o modelo proposto é uma ferramenta valiosa para a predição da ecotoxicidade de fármacos, oferecendo uma abordagem sustentável e confiável para a gestão de contaminantes emergentes no ambiente aquático.

## Agradecimentos

Os autores agradecem, com apreço, à FAPEMIG pelo apoio financeiro.



64º Congresso Brasileiro de Química  
04 a 07 de novembro de 2025  
Belo Horizonte - MG

## Referências

SANGION, A.; GRAMATICA, P. Hazard of Pharmaceuticals for Aquatic Environment: Prioritization by Structural Approaches and Prediction of Ecotoxicity. **Environment International**, 95, 131–143, 2016.

MARTINS, J. P. A.; BARBOSA, E. G.; PASQUALOTO, K. F. M.; FERREIRA, M. M. C. LQTA-QSAR: A New 4D-QSAR Methodology. **J. Chem. Inf. Model.**, 49, 1428–1436, 2009.

ROY, K.; DAS, R. N.; AMBURE, P.; AHER, R. B. Be aware of error measures. Further studies on validation of predictive QSAR models. **Chemometrics and Intelligent Laboratory Systems**, 152, 18–33, 2016.