



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I -- CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA
AMBIENTAL**

LUANA DE MORAIS E SILVA

**FERRAMENTAS *IN SILICO* APLICADAS AO ESTUDO DE ECOTOXICIDADE
DE MICROPOLUENTES ORGÂNICOS PERSISTENTES**

CAMPINA GRANDE

2021

LUANA DE MORAIS E SILVA

**FERRAMENTAS *IN SILICO* APLICADAS AO ESTUDO DE ECOTOXICIDADE
DE MICROPOLUENTES ORGÂNICOS PERSISTENTES**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia Ambiental da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Doutora em Engenharia Ambiental.

Área de concentração: Tecnologias de Tratamento de Água e Resíduos

Orientador: Prof. Dr. Wilton Silva Lopes

Coorientador: Prof. Dr. Marcus Tullius Scotti

CAMPINA GRANDE

2021

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586f Silva, Luana de Morais e.
Ferramentas *in silico* aplicadas ao estudo de ecotoxicidade de micropoluentes orgânicos persistentes [manuscrito] / Luana de Morais e Silva. - 2021.
128 p.

Digitado.
Tese (Doutorado em Engenharia Ambiental) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.
"Orientação : Prof. Dr. Wilton Silva Lopes, Departamento de Engenharia Sanitária e Ambiental - CCT."
1. Micropoluentes orgânicos. 2. Ecotoxicidade. 3. Segurança Química. 4. Quiminformática. I. Título
21. ed. CDD 628.3

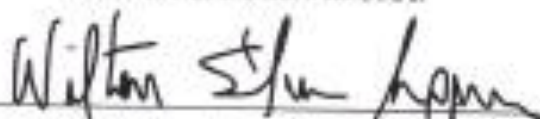
LUANA DE MORAIS E SILVA

Ferramentas *in silico* aplicadas ao estudo de ecotoxicidade de micropoluentes
orgânicos persistentes

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia
Ambiental da Universidade Estadual da Paraíba em cumprimento aos requisitos para a
obtenção do título de Doutora em Engenharia Ambiental.

Aprovado em 30 de março de 2021.

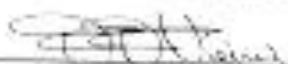
BANCA EXAMINADORA:



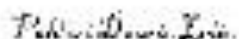
Prof. Dr. Wilton Silva Lopes (Orientador)
Universidade Estadual da Paraíba



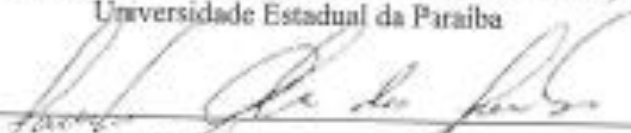
Prof. Dr. Marcus Tullius Scotti (Co-orientador)
Universidade Federal da Paraíba



Prof. Dr. Fernando Fernandes Vieira (Examinador Interno)
Universidade Estadual da Paraíba



Prof. Dr. Valderi Duarte Leite (Examinador Interno)
Universidade Estadual da Paraíba



Prof. Dr. Socrates Golzio Dos Santos (Examinador Externo)
Universidade Federal da Paraíba



Prof. Dr. Ievgen Muratov (Examinador Externo)
University of North Carolina
Universidade Federal da Paraíba

AGRADECIMENTOS

Ao Prof. Dr. Wilton Silva Lopes, pelo aprendizado, confiança, pela oportunidade de realizar este trabalho de doutorado e por sua orientação.

Ao Prof. Dr. Marcus Tullius Scotti, especial agradecimento, pela orientação e pelos ensinamentos essenciais ao desenvolvimento deste trabalho de doutorado.

Aos professores Dr. Eugene Muratov e ao Dr. Vinicius Alves, pelas contribuições e interesse em compartilhar conhecimento.

Aos professores Dr. Eugene Muratov, Dr. Fernando Fernandes, Dr. Sócrates Golzio, Dr. Valderi Leite, por aceitarem participar da banca examinadora, contribuírem para a melhoria do trabalho e para minha formação.

Aos professores do Programa de Pós-Graduação em Ciência e Tecnologia Ambiental (PPGCTA) da Universidade Estadual da Paraíba – UEPB pelo conhecimento transmitido.

Aos funcionários do Programa de Pós-Graduação em Ciência e Tecnologia Ambiental (PPGCTA) da Universidade Estadual da Paraíba – UEPB, pela atenção, profissionalismo e prestatividade.

A todos os colegas do PPGCTA e laboratório de quiminformática da UFPB, pela amizade, aprendizado, ajuda, troca de experiências e apoio.

E a todos que, de que alguma forma, contribuíram para a realização deste trabalho ou que viabilizaram a oportunidade de aprender.

Agradeço também às agências de fomento à pesquisa FAPESQ, CAPES e CNPQ pelo apoio financeiro concedido.

“Aprender é a única coisa que a mente nunca se cansa, nunca tem medo e nunca se arrepende” (Da Vinci)

RESUMO

Micropoluentes orgânicos (MPO), incluem uma ampla gama de substâncias químicas extensivamente utilizadas que ocorrem tipicamente em baixas concentrações na ordem de nanogramas por litro (ng.L^{-1}) a microgramas por litro ($\mu\text{g.L}^{-1}$). Embora ocorram em baixas concentrações, os MPO são continuamente liberados e amplamente detectados no ambiente. E o potencial risco da presença de um grande número de MPO ainda não foram identificadas nem regulamentadas. Para estabelecer limites regulatórios para MPO, pesquisas adicionais sobre respostas biológicas desses compostos são de particular importância. Frente a este cenário, o presente trabalho utilizou métodos preditivos computacionais a fim de fornecer um suporte teórico para analisar parâmetros de interesse no contexto da avaliação de segurança química e desenvolver informações necessárias para preencher as lacunas nos dados para melhorar a compreensão da toxicidade e do potencial perigo dos MPO. Neste estudo abordagens computacionais foram desenvolvidas para identificar as propriedades moleculares relacionadas ao mecanismo de toxicidade aguda de peixes frente a compostos definidos como MoA 1. Para prever a toxicidade de *Daphnia magna* e *Pimephales promelas* para compostos identificados em diferentes fontes de água no Brasil. Para avaliação da segurança química de produtos de transformação de lixiviado de aterro sanitário formado durante o processo Fenton. Os modelos foram desenvolvidos utilizando descritores VolSurf + v. 1.0.7, Dragon v.7.04, RDKit, impressões digitais Morgan (semelhantes a ECFP6), chaves MACCS e impressões digitais Avalon. Para análise dos dados foram empregados métodos de mínimos quadrados parciais (PLS, do inglês, *partial least squares*) e algoritmo Random Forest. De acordo com as melhores práticas de modelagem estes modelos foram estatisticamente robustos, externamente preditivos e caracterizados por um amplo domínio de aplicabilidade estrutural, demonstrando bom potencial preditivo para todos os parâmetros investigados. As informações aqui relatadas foram utilizadas como ferramentas de triagem e avaliação de toxicidade e segurança química de micropoluentes orgânicos.

Palavras-chave: Micropoluentes orgânicos. Ecotoxicidade. Segurança química. Quiminformática.

ABSTRACT

Organic micropollutants (MPO) include a wide range of widely used applied substances that occur in low orders on the order of nanograms per liter (ng.L⁻¹) to micrograms per liter (µg.L⁻¹). Although they occur at low casualties, MPO are continuously released and widely detected in the environment. The potential risk of the presence of many MPO has not yet been identified or regulated. To establish regulatory limits for MPO, additional research on the biological responses of these compounds is of particular importance. Given this scenario, the present work used computational predictive methods to provide theoretical support to analyze parameters of interest in the context of chemical safety assessment and to develop information necessary to fill data gaps to improve understanding of toxicity and potential danger of the MPO. In this study, computational approaches were developed to identify the molecular properties related to the acute toxicity mechanism of fish against compounds defined as MoA 1. To predict the toxicity of *Daphnia magna* and *Pimephales promelas* for compounds identified in different water sources in Brazil. To assess the chemical safety of leachate transformation products from landfill formed during the Fenton process. The models were developed using VolSurf + v descriptors. 1.0.7, Dragon v.7.04, RDKit, Morgan fingerprints (ECFP6-like), MACCS keys and Avalon fingerprints. Partial least squares (PLS, English, partial least squares) and Random Forest algorithm were used for data analysis. According to the best modeling practices, these models were developed and rigorously validated, demonstrating good predictive potential for all investigated parameters. The information reported here was used as a tool for screening and assessing the toxicity and chemical safety of organic micropollutants.

Keyword: Organic micropollutants. Ecotoxicity. Chemical safety. Cheminformatics.

LISTA DE FIGURAS

CAPÍTULO I: INTRODUÇÃO

- Figura 1 - Esquema representando a geração de um modelo de QSAR.....23
- Figura 2 - Fluxograma mostrando as etapas gerais para a geração de modelos QSAR 26

CAPÍTULO II: PREDIÇÃO DE ECOTOXICIDADE DE QUÍMICOS ORGÂNICOS CLASSIFICADOS COMO MOA 1 USANDO ABORDAGEM *IN SILICO*

- Figura 1 - Gráfico de R^2 e Q^2 versus o número de VLs do modelo PLS. O modelo foi selecionado contendo duas variáveis ($R^2 = 0,823$ and $Qcv^2 = 0,793$)49
- Figura 2 - Gráfico PLS dos valores predito vs. experimentais para os dados de toxicidade aguda dos peixes. Os compostos externos são representados por quadrados cinzas e o conjunto de treinamento são os pontos em negrito50
- Figura 3 - Gráfico de escores dos 37 compostos do conjunto de treino (VL1 versus VL2) do modelo PLS. Os pontos em negrito representam os compostos com maiores valores de toxicidade aguda de peixes51
- Figura 4 - Gráfico de escores com 25 compostos do conjunto de validação externa (VL1 versus VL2) do modelo PLS. Os pontos mais escuros representam os compostos com maiores valores de toxicidade aguda de peixes.....51
- Figura 5- Gráfico de peso (VL1 versus VL2) do modelo PLS. Os rótulos dos descritores moleculares (triângulos) referem-se aos símbolos listados na Tabela S3. O quadrado representa a variável dependente, pEC_{50} , o logaritmo negativo da EC_{50} - 50% da concentração efetiva (mol/L) para a toxicidade aguda dos peixes.....54
- Figura 6 - MIFs: as regiões verdes representam uma sonda hidrofóbica em um nível de energia de -0,2 kcal/mol para compostos de maior toxicidade (8, 39, 79) e de menor toxicidade (12, 82, 86).....56

CAPÍTULO III: FERRAMENTAS COMPUTACIONAIS PREDITIVAS PARA AVALIAÇÃO DE ECOTOXICIDADE DE MICROPOLUENTES ORGÂNICOS ENCONTRADOS EM FONTES AQUÁTICAS DO BRASIL

Figura 1 - Gráfico de regressão dos valores de pLC₅₀ predito versus observados para *D. magna* (a e c) e valores observados versus preditos de pLC₅₀ (b e d) para consenso Random Forest (a e b) e software T.E.S.T. (c e d).....79

CAPÍTULO IV: AVALIAÇÃO DA SEGURANÇA QUÍMICA DE PRODUTOS DE TRANSFORMAÇÃO DE LIXIVIADO DE ATERRO SANITÁRIO FORMADO DURANTE O PROCESSO FENTON

Figura 1 - Análise de agrupamento de 154 compostos identificados em lixiviado bruto (A) e 82 compostos identificados em lixiviado tratado (B).....106

Figura 2 – Estrutura química e predições correspondentes para log K_{ow}.115

Figura 3 – Exemplo ilustrativo das estruturas químicas de compostos classificados como biodegradáveis (A) e não biodegradáveis (B).115

Figura 4 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Tetrahymena pyriformis*.....116

Figura 5 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Daphnia magna*.....117

Figura 6 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Pimephales promelas*.....118

Figura 7 - Mapa de calor para visualizar a matriz de correlação de diferentes *endpoints* de lixiviados bruto (A) e tratado (B). A cor verde está associada a compostos não tóxicos/biodegradáveis e vermelhos - a compostos tóxicos/não biodegradáveis.....121

LISTA DE TABELAS

CAPÍTULO II: PREDIÇÃO DE ECOTOXICIDADE DE QUÍMICOS ORGÂNICOS CLASSIFICADOS COMO MOA 1 USANDO ABORDAGEM *IN SILICO*

Tabela 1- Estruturas químicas no formato SMILES e valores experimentais de pEC50 para predição externa e respectivos erros.....	41
Tabela 2 - Estruturas químicas no formato SMILES usada no conjunto de treino aplicado no modelo PLS neste estudo.....	43
Tabela 3 - Descritores moleculares selecionados no modelo QSAR.	53

CAPÍTULO III: FERRAMENTAS COMPUTACIONAIS PREDITIVAS PARA AVALIAÇÃO DE ECOTOXICIDADE DE MICROPOLUENTES ORGÂNICOS ENCONTRADOS EM FONTES AQUÁTICAS DO BRASIL

Tabela 1 - Conjunto de dados de <i>D. magna</i> e <i>P. promelas</i>	73
Tabela 2 - Coeficiente de determinação de validação cruzada R2 de 10 vezes e conjuntos de testes (Rext2) para os modelos Dragon, VolSurf e Consensus Random Forest.....	73
Tabela 3 - O coeficiente de correlação por validação cruzada (Qcv2) e a variância explicada na predição (Qext2) para os modelos Dragon, VolSurf e Consensus Random Forest.....	74
Tabela 4 - Valores experimentais e previstos de pLC50 de <i>D. magna</i> obtidos pelas quatro diferentes abordagens.....	75
Tabela 5 - Matriz de correlação (r) dos valores previstos de LC50 de <i>D. magna</i> pelos quatro métodos utilizados neste estudo para 64 MPO identificados em diferentes matrizes de água no Brasil. a) Valores de LC50 preditos apenas para 54 compostos.....	77
Tabela 6 - Coeficiente de determinação e RMSE dos valores previstos de <i>D. magna</i> pLC50 de 20 compostos de MPO identificados em diferentes matrizes de água no Brasil.....	78
Tabela 7 - Os valores dos parâmetros (em negrito estão dentro da faixa permitida) utilizados pelo critério de validação do modelo, proposto por Tropsha e Roy, de CM e T.E.S.T. modelos para valores de <i>D. magna</i> pLC50.....	79
Tabela 8 - Os MPO potencialmente tóxicos selecionados contra <i>D. magna</i> e <i>P. promelas</i>	82
Tabela 9 - Predição qualitativa de CM e SARpy/IRF e confiabilidade da predição SARpy/IRF. A classificação é baseada nos valores de intervalo de LC50. Tóxico 1, 2, 3 e não-	

tóxico, respectivamente, valores de LC ₅₀ entre 1 e 10 mg/L, 10 e 50 mg/L, 50 e 100 mg/L, e > 100 mg/L. Em negrito estão destacados a combinação de previsão de ambas as abordagens.....	83
Tabela 10 - Matriz de correlação (r) dos valores preditos de LC ₅₀ de <i>D. magna</i> pelos três métodos utilizados neste estudo para 64 MPO identificados em diferentes matrizes de água no Brasil. a) CM previu valores de LC ₅₀ para 46 compostos, b) ToxRead para 45 e c) T.E.S.T. para 57 compostos.....	85
Tabela 11 - Valores experimentais e previstos de pLC ₅₀ de <i>P. promelas</i> obtidos pelas três diferentes abordagens.....	86
Tabela 12 - Coeficiente de determinação e o RMSE dos valores estimados de LC ₅₀ de <i>D. magna</i> de 20 compostos de MPO identificados em diferentes fontes de água no Brasil.....	88

CAPÍTULO IV: AVALIAÇÃO DA SEGURANÇA QUÍMICA DE PRODUTOS DE TRANSFORMAÇÃO DE LIXIVIADO DE ATERRO SANITÁRIO FORMADO DURANTE O PROCESSO FENTON

Tabela 1 - Exemplos de estruturas químicas presentes nos principais agrupamentos identificados para a amostra de lixiviado de aterro sanitário bruto.....	107
Tabela 2 - Exemplos de estruturas químicas presentes nos principais agrupamentos identificados para a amostra de lixiviado tratado por Fenton.....	109
Tabela 3 - Características estatísticas dos modelos de QSAR, MuDRA e modelos de consenso para as predições das atividades selecionadas.....	110
Tabela 4 - Resultados de QSAR, MuDRA e previsões de consenso para as atividades selecionadas.....	112

LISTA DE SIGLAS

CAS	– <i>Chemical Abstract Services</i>
CE ₅₀	– Concentração Efetiva da concentração da amostra que causa um efeito agudo a 50% dos organismos
CL ₅₀	– Concentração Letal da concentração da amostra que causa mortalidade de 50% dos organismos
DA	– Domínio de aplicabilidade
KNIME	– Konstanz Information Miner
MDL	– Molecular Design Ltd
MIF	– Molecular Interaction Field
MoA	– Modo de Ação
OCDE	– Organização para Cooperação e Desenvolvimento Econômico
PLS	– <i>Partial Least Squares</i>
Q ²	– Coeficiente de correlação de validação cruzada externa
(Q)SAR	– <i>(Quantitative) Structure-Activity Relationships</i>
R ²	– Coeficiente de correlação cruzada
REACH	– <i>Registration, Evaluation and Authorization of Chemicals</i>
RF	– <i>Random forest</i>
RMSE	– <i>Root Mean Squared of Errors</i>
SDEC	– <i>Standard Deviation Error in Calculation</i>
SDEP	– <i>Standard Deviation Error in Prediction</i>
SMILES	– <i>Simplified Molecular Input Line Entry System</i>
TD ₅₀	– <i>Dose for 50% toxic effect</i>
US EPA	– U.S. Environmental Protection Agency
VIP	– <i>Variable Influence on Projection</i>
VLs	– Variáveis Latentes

SUMÁRIO

CAPÍTULO I - INTRODUÇÃO	16
1 MICROPOLUENTES ORGÂNICOS	16
2 AVALIAÇÃO DE RISCO DE PRODUTOS QUÍMICOS	19
3 REGULAMENTAÇÃO DE PRODUTOS QUÍMICOS	20
4 FERRAMENTAS COMPUTACIONAIS PARA PREDIÇÃO DE ECOTOXICIDADE	22
5 OBJETIVO	27
REFERÊNCIAS	28
CAPÍTULO II - PREDIÇÃO DE ECOTOXICIDADE DE QUÍMICOS ORGÂNICOS CLASSIFICADOS COMO MOA 1 USANDO ABORDAGEM <i>IN SILICO</i>	34
RESUMO	34
ABSTRACT	35
1 INTRODUÇÃO	36
2 MATERIAL E MÉTODOS	38
2.1 Conjunto de dados	38
2.2 Descritores moleculares	38
2.3 Análise estatística e validação do modelo	39
2.4 Domínio de Aplicabilidade (DA)	39
3 RESULTADOS E DISCUSSÕES	40
3.1 Geração de Modelo	40
3.2 Modelo de refinamento e interpretação química	50
3.3 Interpretação do modelo	52
4 CONCLUSÃO	56
REFERÊNCIAS	58
CAPÍTULO III - FERRAMENTAS COMPUTACIONAIS PREDITIVAS PARA AVALIAÇÃO DE ECOTOXICIDADE DE MICROPOLUENTES ORGÂNICOS ENCONTRADOS EM FONTES AQUÁTICAS DO BRASIL	65
RESUMO	65
ABSTRACT	66

1 INTRODUÇÃO	67
2 MATERIAL E MÉTODOS	70
2.1 Conjuntos de dados	70
2.2 Desenvolvimento e validação de dados	71
3 RESULTADOS E DISCUSSÕES	72
3.1 Desenvolvimento e Validação dos Modelos.....	73
3.2 Desempenho de diferentes abordagens.....	74
3.2.1 Modelagem de MPO para <i>Daphnia magna</i>	75
3.2.2 Modelagem de MPO para <i>Pimephales promelas</i>	81
4 CONCLUSÃO	88
REFERÊNCIAS	89
CAPÍTULO IV - AVALIAÇÃO DA SEGURANÇA QUÍMICA DE PRODUTOS DE TRANSFORMAÇÃO DE LIXIVIADO DE ATERRO SANITÁRIO FORMADO DURANTE O PROCESSO FENTON	96
RESUMO	96
ABSTRACT	97
1 INTRODUÇÃO	98
2 MATERIAL E MÉTODOS	100
2.1 Análise Experimental.....	100
2.1.1 Amostras de lixiviado e processo Fenton	100
2.1.2 Métodos analíticos para a elucidação de estruturas químicas	100
2.2 Abordagem quiminformatica	101
2.2.1 Conjunto de dados	101
2.2.2 Descritores moleculares	101
2.2.3 Geração dos modelos de QSAR	102
2.2.4 Análise de agrupamento químico	103
2.2.5 Análise estatística	103
3 RESULTADOS E DISCUSSÕES	104
3.1 Identificação de compostos em lixiviado de aterro sanitário e seus TPs formados durante o processo Fenton	104
3.2 Análise de agrupamento hierárquico	104
3.3 Modelos <i>in silico</i>	108

3.3.1 Log K_{ow}	114
3.3.2 Biodegradabilidade	114
3.3.3 <i>Tetrahymena pyriformis</i>	116
3.3.4 <i>Daphnia magna</i>	116
3.3.5 <i>Pimephales promelas</i>	117
3.3.6 AhR e ARE	119
3.3.7 AMES	119
3.3.8 Análise de similaridade das atividades	119
4 CONCLUSÃO	122
REFERÊNCIAS	123

CAPÍTULO I

CAPÍTULO I - INTRODUÇÃO

1 MICROPOLUENTES ORGÂNICOS

Os produtos químicos são indispensáveis, por exemplo, para a saúde, alimentação e produção industrial, e para manter nosso padrão de vida, mas também para os ecossistemas. No entanto, como uma consequência do seu uso, dezenas de milhares de diferentes químicos são emitidos para o meio ambiente, resultando em misturas muito complexas e freqüentemente desconhecidas de produtos químicos, provocando o aumento da poluição. O que representa uma crescente preocupação ambiental, visto que podem provocar toxicidade e efeitos adversos à saúde humana e ao ecossistema (GAVRILESCU et al., 2015; HAMID; LI; GRACE, 2018; PETRIE; BARDEN; KASPRZYK-HORDERN, 2015).

Micropoluentes orgânicos (MPO), também chamados de contaminantes emergentes, incluem uma ampla gama de substâncias químicas extensivamente utilizadas desde residências a aplicações industriais, tais como produtos farmacêuticos e produtos de cuidados pessoais (BIELEN et al., 2017; MANDARIC et al., 2017), desreguladores endócrinos (MELO; BRITO, 2014), pesticidas (MUNZ et al., 2017), tensoativos (FEDEILA et al., 2018), nanomateriais (SANTHOSH et al., 2016), metabólitos e seus produtos de transformação que ocorrem tipicamente em baixas concentrações na ordem de nanogramas por litro (ng.L^{-1}) a microgramas por litro ($\mu\text{g.L}^{-1}$) (DEEB et al., 2017). A principal fonte antrópica para o lançamento de MPO para o meio ambiente tem sido principalmente atribuída às estações de tratamento de águas residuais urbanas (ETARs), devido ao uso de técnicas de tratamentos convencionais secundários (por exemplo, lodo ativado) e terciários (como filtração e desinfecção) não eficazes na remoção da maioria dos MPOs que entram em ETARs (BUI et al., 2016; GROS et al., 2017; LUO et al., 2014). Outras fontes de MPO incluem: (i) águas residuais industriais, (ii) escoamento agrícola, pecuária (BARBOSA et al., 2016) e aquicultura (TAVAKOL et al., 2017), (iii) lixiviado de aterros sanitários; e (iv) efluentes hospitalares, dos quais micropoluentes podem seguir muitos caminhos (WANG; WANG; YANG, 2018). Conseqüentemente, isso leva à detecção de muitos micropoluentes orgânicos em diferentes fontes aquáticas, como águas superficiais (NEALE et al., 2017), água

subterrâneas (AN et al., 2018), águas oceânicas (NÖDLER; VOUTSA; LICHA, 2014) e até mesmo água potável (YANG et al., 2017).

Embora ocorram em baixas concentrações, os MPO são continuamente liberados no meio aquático com propriedades físico-químicas particulares, muitas vezes reconhecidas por propriedades indesejáveis nas quais podem exercer efeitos nocivos sobre o ecossistema e a saúde humana (ARNOT; MACKAY, 2008). Dentre eles compostos resistentes à degradação e altamente persistentes em meios aquosos, que apresentam um risco potencial de toxicidade quando no meio ambiente (DESBIOLLES et al., 2018; RIVERA-UTRILLA et al., 2013). Químicos estrogênicos os quais podem interferir significativamente na função do sistema endócrino (BARREIROS et al., 2016), associado às funções de desenvolvimento e reprodução, particularmente em organismos aquáticos (KOBAYASHI et al., 2017). A ocorrência e a diversidade de antibióticos no ambiente aquático têm provocado o desenvolvimento e disseminação de linhagens de bactérias resistentes aos antibióticos (JIA et al., 2018; TOOLARAM et al., 2016). Recentes estudos sobre o acúmulo de poluentes orgânicos persistentes em diversos animais selvagens aquáticos e terrestres, tem indicando potencial de bioacumulação dependente do habitat entre diferentes espécies e potencial de transferência materna (LIU et al., 2018). Da mesma forma, os pesticidas, uma das categorias mais comuns de poluentes, formados por um vasta gama de compostos e produtos químicos podem ser persistentes em água ou bioacumuláveis na biota (CCANCCAPA et al., 2016; HEYS et al., 2016). Um dos principais problemas relacionados ao uso destes compostos é que eles não são totalmente seletivos, representando uma ameaça ao funcionamento normal dos ecossistemas aquáticos e terrestres (MANSANO et al., 2016).

Ainda que amplamente detectada no ambiente, o potencial risco da presença de um grande número de MPO ainda não foram identificados nem regulamentados. Além disso, a quantidade e diversidade impedem uma avaliação de risco abrangente, principalmente devido à complexidade de sua análise experimental, que geralmente é demorada e cara (KHALEEL et al., 2019; KÜMMERER, 2019)(KHALEEL et al., 2019; KÜMMERER et al., 2019). Portanto, a avaliação experimental não é adequada para a triagem sob demanda da ampla variedade de MPO, o que dificulta o conhecimento sobre os efeitos desses compostos em efluentes complexos.

Ações inovadoras nessa direção têm sido desenvolvidas por agências reguladoras como a Agência de Proteção Ambiental dos Estados Unidos (do inglês, *United States Environmental Protection Agency – U.S. EPA*) e Agência Europeia das Substâncias Químicas a partir do regulamento de Registro, Avaliação, Autorização e Restrição de Produtos Químicos (do inglês, *Registration, Evaluation, Authorisation and Restriction of Chemicals - REACH*) (BARBOSA *et al.*, 2016) orientando para a tomada de decisões e desenvolvimento de políticas que promova o uso de métodos alternativos para a caracterização de risco, avaliação de risco e gestão da ecotoxicidade a fim de reduzir o número de testes em animais (MONTAGNER, *et al.*, 2017; OECD, 2014 (PORCELLI *et al.*, 2008)).

Métodos computacionais baseados em estudo matemático de relações quantitativas entre estrutura química e a atividade/propriedade de um composto [do inglês (*Quantitative Structure-Activity Relationships - (Q)SARs*)] tem desempenhado um papel importante no campo da química medicinal (CAPUZZI *et al.*, 2018), design de drogas (ALVES *et al.*, 2016) e toxicologia (GRAMATICA *et al.*, 2012), atualmente tornando-se cada vez mais aplicada nas ciências ambientais como ferramentas de triagem para classificar e priorizar produtos químicos que podem apresentar impacto adverso na saúde humana e ambiental (GRAMATICA; PAPA; SANGION, 2018). O estudo *in silico* baseado em QSAR é derivado da aplicação de ferramentas matemáticas e estatísticas, resultando em uma correlação entre propriedades físico-químicas e atividades biológicas, numericamente codificadas por um ou mais descritores moleculares que são usados para prever as propriedades de interesse (CONSONNI; BALLABIO; TODESCHINI, 2009; CONSONNI; TODESCHINI; PAVAN, 2002; DEARDEN; ROTUREAU; FAYET, 2013; GASTEIGER, 2016). A aplicação de tais métodos é uma alternativa útil para lidar com grandes quantidades de dados, avaliar as propriedades das substâncias, seus processos determinantes de destino e prever seus efeitos, reduzindo experimentos caros e demorados e mostrando um forte potencial para reduzir o uso de animais utilizados em ensaios experimentais (DE MORAIS E SILVA *et al.*, 2018; ECHA, 2014; NICOLAS *et al.*, 2018). Em um contexto de tomada de decisão, a proteção abrangente da saúde pública e do meio ambiente com o uso de métodos preditivos computacionais apoia os esforços contínuos para desenvolver informações a fim de preencher as lacunas nos dados de eco(toxicidade), necessário para melhorar a compreensão

da avaliação dos perigos dos produtos químicos conforme exigido por várias agências governamentais em todo o mundo, incluindo a U.S. EPA e regulamentos adotados pela UE através do REACH (EBELE; ABOU-ELWAFABDALLAH; HARRAD, 2017; TODESCHINI, 2009; TOXICITY, 2008).

2 AVALIAÇÃO DE RISCO DE PRODUTOS QUÍMICOS

A avaliação do potencial risco de substâncias química tem sido cada vez mais abordada no exame de problemas ambientais e aplicada para apoiar o processo de tomada de decisão de gerenciamento de produtos químicos, novos e existentes. Este processo de controle químico tem como objetivo examinar as exposições resultantes de descargas e/ou liberações de produtos químicos e caracterizar o efeito de tais emissões à saúde humana e ao meio ambiente (ECB, 2003; FIRESTONE et al., 2010)

A avaliação da exposição combinada com informações detalhadas sobre dados toxicológicos inerentes a estes produtos químicos são ferramentas fundamentais as boas avaliações de riscos e necessárias para prever a probabilidade, a natureza e a magnitude dos efeitos adversos que podem acometer a saúde humana e o meio ambiental (NORTON et al., 1992; RAPPAPORT, 2011).

O processo de avaliação de risco consiste em quatro etapas principais, que incluem:

(1) identificação de perigos: identificação dos efeitos adversos os quais uma substância é capaz de causar. O objetivo é coletar e avaliar dados sobre os efeitos à saúde (para seres humanos e meio ambientes) que pode ser produzido pela exposição a um produto químico, bem como a caracterização do comportamento de um produto químico dentro do corpo e no ambiente. O método de identificação de perigo usa dados de animais (testes toxicológicos), dados de humanos (geralmente, avaliação epidemiológica) outros dados incluem relação estrutura-atividade ou dados *in vitro*.

(2) avaliação dose/resposta: depois de identificar o perigo, esta avaliação examina as relações quantitativas entre concentração/doses e os efeitos de um produto químico ao quais populações ou compartimentos ambientais estão ou podem estar expostos. A exposição pode ser avaliada medindo as concentrações de exposição ou aplicando modelos de exposição, que

estimam as emissões, caminhos e taxas de movimento de uma substância e sua transformação ou degradação.

(3) avaliação da exposição: é o processo para estimar qualitativamente e /ou quantitativamente a magnitude, frequência, extensão, caráter e duração de exposições de um agente no ambiente ou estimar exposições futuras para um agente. Dados podem ser obtidos de estudos experimentais de laboratório, de estudos de campo e epidemiológicos, mas também de metodologias alternativas, incluindo testes *in vitro* e estimativas *in-silico*, tais como relações estrutura-atividade (QSARs), que pode facilitar a previsão de propriedade toxicológicas de produtos químicos, ou onde nenhum teste tenha sido realizado para estabelecer o perfil toxicológico de um novo produto químico. A saída desta etapa é a identificação de um nível seguro sob o qual não se espera que ocorram efeitos adversos.

(4) caracterização de risco: a caracterização do risco é a fase final, este integra dois componentes principais: estimativa do risco e descrição do risco. "Estimativa de risco" combina perfis de exposição e efeitos de exposição. A "descrição do risco" fornece informações importantes para interpretar os resultados do risco e identifica um nível de efeitos prejudiciais dos estressores nos alvos de interesse.

Nos Estados Unidos, União Europeia e em alguns outros países, as metodologias de avaliação de risco por produtos nocivos à saúde humana e ao ecossistema são apoiadas por legislação exigente que inclui a aplicação de procedimentos bem definidos para auxiliar na resposta regulatória apropriada para um risco potencial à saúde ou ao meio ambiente (enHEALTH, 2012).

3 REGULAMENTAÇÃO DE PRODUTOS QUÍMICOS

Os programas de regulamentação destinados a avaliar e gerir os riscos dos produtos químicos requerem informações sobre uma vasta gama de propriedades e efeitos químicos, incluindo propriedades físico-químicas e ambientais, bem como efeitos sobre a saúde humana e espécies ambientais (WORTH et al., 2010). Na União Europeia (UE), por exemplo, o quadro legislativo no contexto do controlo de produto químico e da gestão de riscos é descrito como mais importante e rígidos nos últimos 20 anos. O Regulamento (EC) N° 1907/2006 do Parlamento Europeu e do Conselho relativo ao registo, avaliação,

autorização e restrição de substâncias químicas (REACH) entrou em vigor em 2007, com principal objetivo de melhorar a proteção aos seres humanos, a vida selvagem e ao meio ambiente contra os potenciais riscos aos quais as substâncias químicas podem representar (ECHA, 2008).

O regulamento busca estabelecer procedimentos para avaliação de informações sobre as propriedades e perigos para todas as substâncias químicas; não apenas as usadas em processos industriais, mas também as que fazem parte da nossa vida diária, como, por exemplo, presentes em produtos de limpeza e tintas, em artigos de vestuário, mobiliário ou aparelhos elétricos, que sejam fabricadas ou importadas dentro da UE ou de países externos, em quantidades superiores a uma tonelada por ano (REACH, 2014).

O REACH baseia-se no conceito de que a indústria tem o papel de garantir que os produtos químicos que fabrica e coloca no mercado da UE não afetam negativamente a saúde humana ou o ambiente. Resumidamente, quatro ações principais regem o REACH:

- Registro de todas as substâncias fabricadas ou importadas na UE em quantidades superiores a 1 tonelada / ano. Sendo necessária uma avaliação de segurança química que inclui a classificação de perigo e a avaliação se a substância é persistente, bioacumulável ou tóxica. Restringindo as substâncias de grande preocupação que podem estar ligadas a câncer, irritação da pele, doenças respiratórias, distúrbios da visão, asma, desregulação endócrina, entre outros.

- Avaliação de dossiês de registro; propostas de ensaio a fim de evitar testes desnecessário com animais vertebrados e potenciais riscos associados a saúde humana e ambiental, identificar substâncias de maior preocupação.

- Autorização para substâncias que suscitem elevada preocupação sejam progressivamente substituídas por substâncias ou tecnologias menos perigosas.

- Restrição utilizadas para limitar ou proibir a produção, a colocação no mercado (incluindo as importações) ou a utilização de uma substância.

Ao longo da implementação do regulamento REACH, algumas preocupações surgiram diante a crescente necessidade por realizar testes de laboratório para avaliação de risco das substâncias a serem registradas. Para reduzir e/ou evitar o uso de pesquisas e testes em animais, bem como potenciais custos, novas abordagens e métodos alternativos de avaliação dos perigos das substâncias foram sugeridos pelo regulamento REACH, como, por

exemplo, métodos *in vitro*, abordagens baseadas na similaridade (métodos comparativos, agrupamento químico) e em particular modelos de relações quantitativas estrutura-atividade (QSAR).

Os cenários previstos pelo REACH para métodos alternativos de teste geraram interesse na comunidade científica. No campo das metodologias baseadas modelos computacionais, isso também é comprovado pelo grande número de projetos que visam desenvolver modelos matemáticos e projetar sistemas de informação que possam atender aos requisitos regulatórios (EUROPEAN COMMISSION ENVIRONMENT DIRECTORATE GENERAL, 2007)

4 FERRAMENTAS COMPUTACIONAIS PARA PREDIÇÃO DE ECOTOXICIDADE

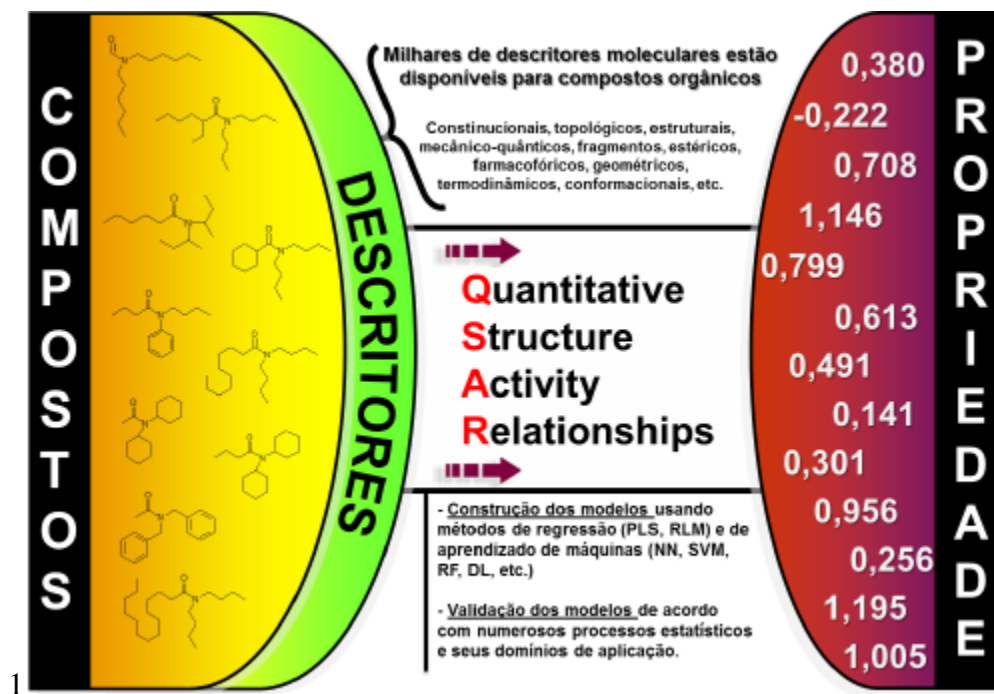
Ferramentas computacionais são cada vez mais empregadas na maioria dos campos da pesquisa científica. O uso de computador e tecnologias de informações aplicados para resolver problemas químicos resultou na ciência denominada Quiminformática, com principal aplicação associado ao armazenamento, indexação e busca de informações relacionadas aos compostos químicos. A quiminformática, inicialmente foi desenvolvida para lidar com problemas químicos, com particular ênfase na descoberta de novas drogas. E posteriormente, devido ao progresso da ciência e tecnologia computacional, permitiu inúmeros avanços em várias outras áreas de conhecimento, como biologia, física, bioquímica, estatísticas, matemática e ciência da computação (BROWN, 2009; GASTEIGER, 2016).

Dentre as aplicações da quiminformática a modelagem quantitativa de relação estrutura-atividade (QSAR, do inglês: *Quantitative Structure-Activity Relationships*) é uma importante ferramenta para as ciências ambientais e biológicas, destinada a fornecer estimativas de muitos resultados de testes laboratoriais antes que os testes sejam conduzidos. A relação estrutura-atividade (SARs) são modelos que relacionam as características estruturais de um composto à sua atividade biológica (HANSH, 1964). E abordagem *in silico* podem ser desenvolvida para prever propriedades físico-químicas, biológicas ou ambientais de um determinado produto químico com base na sua estrutura química, bem

como compreender e racionalizar os mecanismos de ação dentro de uma série de produtos químicos (Figura 1) (SIPPL, 2010).

A análise da relação entre as estruturas químicas e atividade/propriedade é representada por descritores moleculares que são obtidos através de um procedimento lógico e matemático que transforma a informação química codificada dentro de uma representação simbólica de uma molécula em um número útil ou o resultado de algum experimento padronizado (TODESCHINI, 2009). Com base nos descritores moleculares calculados para um conjunto de compostos e considerando as respectivas atividades biológicas ou propriedades físico-químicas, pode-se estabelecer um QSAR. A base conceitual dos QSARs é o princípio da similaridade que afirma que moléculas com estruturas similares provavelmente podem estar associado a atividades/propriedades semelhantes.

Figura 1 - Esquema representando a geração de um modelo de QSAR.



Fonte: Alves, 2017; Tropsha, 2010.

Muitos descritores moleculares têm sido propostos e derivados de diferentes teorias e abordagens (TODESCHINI e CONSONNI, 2000). Quanto às representações estruturais, os descritores moleculares podem ser classificados de acordo com as dimensões equivalentes

a diferentes níveis de “complexidade”. De acordo com a informação química codificada os principais descritores são:

- Descritores 1D são calculados a partir da fórmula química, os descritores derivados dessa representação são tipicamente usados apenas para o conhecimento parcial da estrutura em relação aos grupos funcionais e fragmentos. A presença de atividade biológica relacionada a uma subestrutura é chamada de alerta estrutural;

- Descritores 2D são baseados na teoria dos grafos, comumente conhecido como representação topológica. Estes descritores são definidos principalmente através de índices topológicos e de conectividade;

- Descritores 3D são derivados das representações geométricas das moléculas e codificam informações não só da natureza e conectividade dos átomos, mas também a configuração espacial global dos átomos da molécula, como: tamanho, forma, superfície, volume de uma conformação estudada da molécula. Esta representação da molécula é definida como representação geométrica.

Diferentes descritores são diferentes maneiras de visualizar uma molécula. A seleção de descritores deve ser realizada aplicando-se abordagens matemáticas para maximizar o poder preditivo do modelo QSAR, já que o principal papel do modelo é sua capacidade preditiva. A fim de garantir a capacidade preditiva do modelo, que produza dados confiáveis, a OECD definiu um conjunto de critérios, bem como procedimentos a serem adotados na realização de um modelo QSAR (“OECD”, 2014):

1) Parâmetro definido - diz respeito às atividades e propriedades das moléculas que possa ser medido e modelado. A intenção é garantir transparência dos dados experimentais utilizados por um determinado modelo.

2) Um algoritmo inequívoco – o modelo deve ser claramente definido, fácil e continuamente aplicável de tal forma que os cálculos para a previsão do *endpoints* possam ser reproduzidos. É necessária uma descrição clara do conjunto de dados, geração de descritores moleculares, procedimento de modelagem e métodos e parâmetros estatísticos utilizados. Essas informações são úteis para estabelecer de forma independente o desempenho e a reprodutibilidade das previsões de um determinado modelo.;

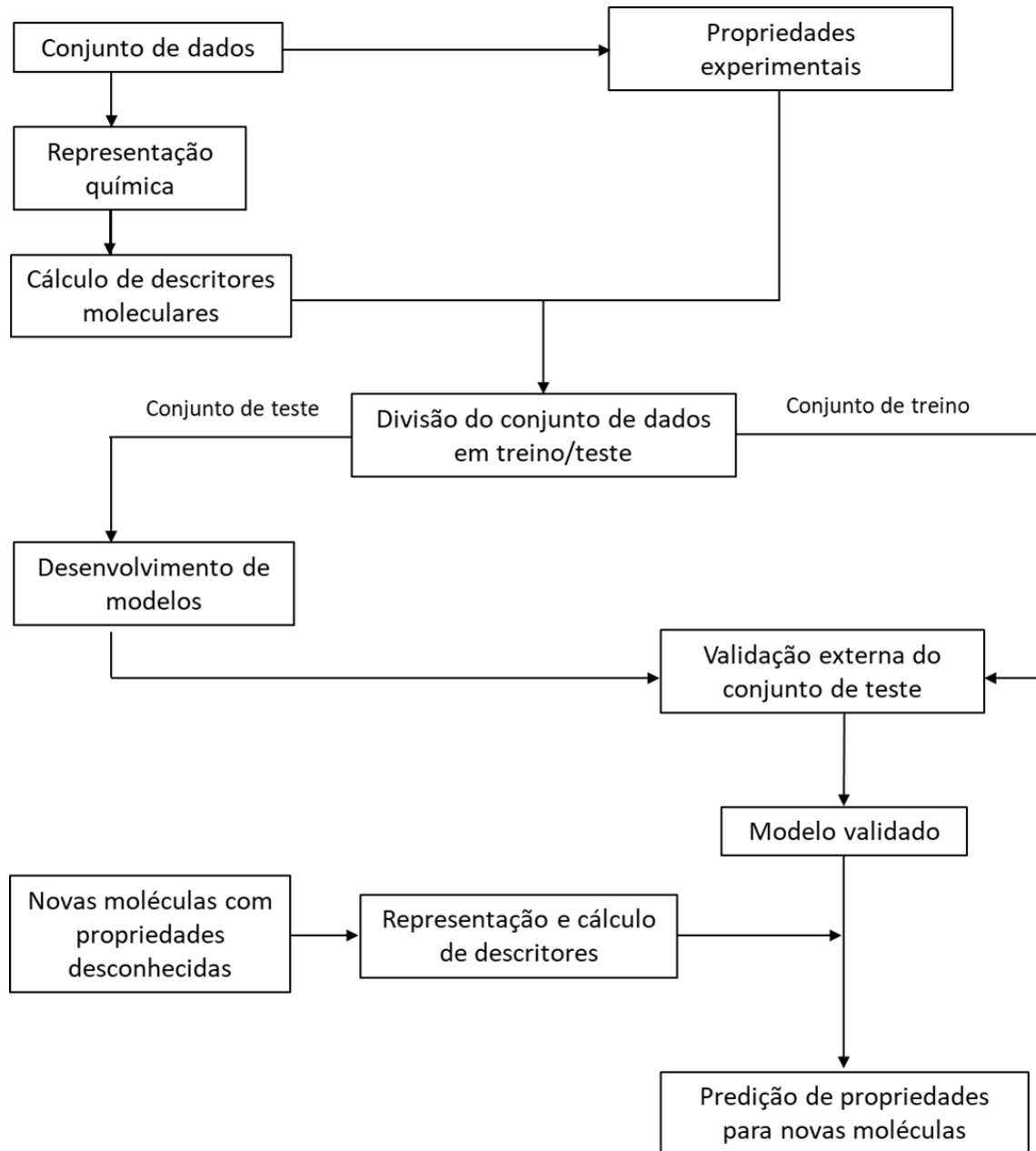
3) Domínio de aplicabilidade definido - é necessário definir um domínio químico baseado nas moléculas de treinamento em que se pressupõe que o modelo forneça previsões precisas (confiáveis);

4) Medidas apropriadas de boa qualidade de ajuste, robustez e previsibilidade - o modelo precisa ser submetido a procedimentos de validação apropriados para estimar não apenas sua capacidade de ajustar os dados no conjunto de treinamento, mas também sua precisão para predição de propriedades para novos produtos químicos;

5) Um mecanismo de interpretação, se possível - a intenção deste quinto princípio é promover a interpretação dos descritores gerado no modelo em relação ao ponto final que o modelo pretendia prever.

O modelo final construído a partir dos melhores ou mais eficazes parâmetros, então, sofrerá validado usando conjuntos de teste para garantir que o modelo é apropriado e útil. O processo completo de construção de um modelo de (Q)SAR é representado na Figura 2 como um fluxograma.

Figura 2 - Fluxograma mostrando as etapas gerais para a geração de modelos QSAR



5 OBJETIVO

Objetivo geral

O objetivo principal do presente trabalho foi desenvolver e aplicar modelos quiminformáticos para avaliação e classificação do perfil (eco)toxicológico de micropoluentes orgânicos.

Objetivo específico

Mais especificamente:

- Aplicar modelo QSAR usando um conjunto de descritores moleculares gerados a partir dos valores de toxicidade aguda de peixes para compostos definidos como MoA1;
- Analisar ecotoxicidade de MPO identificados em diferentes fontes de água no Brasil para crustáceo de água doce (*Daphnia magna*) e peixe (*Pimephales promelas*) através de métodos computacionais;
- Avaliação de segurança química de compostos identificados no lixiviado de aterro sanitário bruto e seus produtos de transformação formados pelo processo Fenton.

REFERÊNCIAS

- ALVES, V. M. et al. Alarms about structural alerts. **Green Chemistry**, v. 18, n. 16, p. 4348–4360, 2016.
- AN, X. et al. Treating organic cyanide-containing groundwater by immobilization of a nitrile-degrading bacterium with a biofilm-forming bacterium using fluidized bed reactors. **Environmental Pollution**, v. 237, p. 908–916, 2018.
- ARNOT, J. A.; MACKAY, D. Policies for chemical hazard and risk priority setting: Can persistence, bioaccumulation, toxicity, and quantity information be combined? **Environmental Science and Technology**, v. 42, n. 13, p. 4648–4654, 2008.
- BARBOSA, M. O. et al. Occurrence and removal of organic micropollutants: An overview of the watch list of EU Decision 2015/495. **Water Research**, v. 94, p. 257–279, 2016.
- BARREIROS, L. et al. Analysis of 17- β -estradiol and 17- α -ethinylestradiol in biological and environmental matrices - A review. **Microchemical Journal**, v. 126, p. 243–262, 2016.
- BIELEN, A. et al. Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. **Water Research**, v. 126, p. 79–87, 2017.
- BROWN, N. Chemoinformatics—An Introduction for Computer Scientists. **ACM Computing Surveys**, v. 41, n. 2, p. 1–38, 2009.
- BUI, X. T. et al. Multicriteria assessment of advanced treatment technologies for micropollutants removal at large-scale applications. **Science of the Total Environment**, v. 563–564, p. 1050–1067, 2016.
- CAPUZZI, S. J. et al. Computer-Aided Discovery and Characterization of Novel Ebola Virus Inhibitors. **Journal of Medicinal Chemistry**, v. 61, n. 8, p. 3582–3594, 2018.
- CCANCCAPA, A. et al. Pesticides in the Ebro River basin: Occurrence and risk assessment. **Environmental Pollution**, v. 211, n. APRIL, p. 414–424, 2016.
- CONSONNI, V.; BALLABIO, D.; TODESCHINI, R. Comments on the definition of the Q2 parameter for QSAR validation. **Journal of Chemical Information and Modeling**, v. 49, n. 7, p. 1669–1678, 2009.
- CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D

molecular descriptors. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 3, p. 682–692, 2002.

DE MORAIS E SILVA, L. et al. Predictive ecotoxicity of MoA 1 of organic chemicals using in silico approaches. **Ecotoxicology and Environmental Safety**, v. 153, n. September 2017, p. 151–159, 2018.

DEARDEN, J. C.; ROTUREAU, P.; FAYET, G. QSPR prediction of physico-chemical properties for REACH. **SAR and QSAR in Environmental Research**, v. 24, n. 4, p. 279–318, 2013.

DEEB, A. A. et al. Suspect screening of micropollutants and their transformation products in advanced wastewater treatment. **Science of the Total Environment**, v. 601–602, p. 1247–1253, 2017.

DESBIOLLES, F. et al. Occurrence and ecotoxicological assessment of pharmaceuticals: Is there a risk for the Mediterranean aquatic environment? **Science of the Total Environment**, v. 639, p. 1334–1348, 2018.

EBELE, A. J.; ABOU-ELWAFI ABDALLAH, M.; HARRAD, S. Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment. **Emerging Contaminants**, v. 3, n. 1, p. 1–16, mar. 2017.

ECB. Document on Risk Assessment. **Technical Guidance Document on Risk Assessment Part II**, p. 337, 2003.

ECHA. **Guidance on information requirements and chemical safety assessment Chapter R . 17 : Estimation of exposure from articles**. [s.l: s.n.].

ECHA. **The Use of Alternatives to Testing on Animals for the REACH Regulation**. [s.l: s.n.]. v. 117

EUROPEAN COMMISSION ENVIRONMENT DIRECTORATE GENERAL. OECD Environment Health and Safety Publications Series on Testing and Assessment No . 69 GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS Environment Directorate. n. February, p. 1–154, 2007.

FEDEILA, M. et al. Biodegradation of anionic surfactants by *Alcaligenes faecalis*, *Enterobacter cloacae* and *Serratia marcescens* strains isolated from industrial wastewater. **Ecotoxicology and Environmental Safety**, v. 163, n. August, p. 629–635, 2018.

FIRESTONE, M. et al. The U.S. environmental protection agency strategic plan for evaluating the toxicity of chemicals. **Journal of Toxicology and Environmental Health - Part B: Critical Reviews**, v. 13, n. 2–4, p. 139–162, 2010.

GASTEIGER, J. Chemoinformatics: Achievements and challenges, a personal view. **Molecules**, v. 21, n. 2, 2016.

GAVRILESCU, M. et al. Emerging pollutants in the environment: Present and future challenges in biomonitoring, ecological risks and bioremediation. **New Biotechnology**, v. 32, n. 1, p. 147–156, 2015.

GRAMATICA, P. et al. QSAR modeling is not “Push a button and find a correlation”: A case study of toxicity of (Benzo-)triazoles on Algae. **Molecular Informatics**, v. 31, n. 11–12, p. 817–835, 2012.

GRAMATICA, P.; PAPA, E.; SANGION, A. QSAR modeling of cumulative environmental end-points for the prioritization of hazardous chemicals. **Environmental Science: Processes and Impacts**, v. 20, n. 1, p. 38–47, 2018.

GROS, M. et al. Screening and prioritization of micropollutants in wastewaters from on-site sewage treatment facilities. **Journal of Hazardous Materials**, v. 328, p. 37–45, 2017.

HAMID, H.; LI, L. Y.; GRACE, J. R. Review of the fate and transformation of per- and polyfluoroalkyl substances (PFASs) in landfills. **Environmental Pollution**, v. 235, p. 74–84, 2018.

HEYS, K. A. et al. Risk assessment of environmental mixture effects. **RSC Advances**, v. 6, n. 53, p. 47844–47857, 2016.

JIA, J. et al. Occurrence and distribution of antibiotics and antibiotic resistance genes in Ba River, China. **Science of the Total Environment**, v. 642, p. 1136–1144, 2018.

KHALEEL, N. D. H. et al. Studying the fate of the drug Chlorprothixene and its photo transformation products in the aquatic environment: Identification, assessment and priority setting by application of a combination of experiments and various in silico assessments. **Water Research**, v. 149, p. 467–476, 2019.

KOBAYASHI, T. et al. Estrogen alters gonadal soma-derived factor (Gsdf)/Foxl2 expression levels in the testes associated with testis-ova differentiation in adult medaka, *Oryzias latipes*. **Aquatic Toxicology**, v. 191, n. May, p. 209–218, 2017.

KÜMMERER, K. From a problem to a business opportunity-design of pharmaceuticals for

environmental biodegradability. **Sustainable Chemistry and Pharmacy**, v. 12, n. January, p. 10–13, 2019.

KÜMMERER, K. et al. Reducing aquatic micropollutants – Increasing the focus on input prevention and integrated emission management. **Science of the Total Environment**, v. 652, p. 836–850, 2019.

LIU, Y. et al. Mechanisms involved in the impact of engineered nanomaterials on the joint toxicity with environmental pollutants. **Ecotoxicology and Environmental Safety**, v. 162, n. May, p. 92–102, 2018.

LUO, Y. et al. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. **Science of the Total Environment**, v. 473–474, p. 619–641, 2014.

MANDARIC, L. et al. Contamination sources and distribution patterns of pharmaceuticals and personal care products in Alpine rivers strongly affected by tourism. **Science of the Total Environment**, v. 590–591, p. 484–494, 2017.

MANSANO, A. S. et al. Effects of diuron and carbofuran pesticides in their pure and commercial forms on *Paramecium caudatum*: The use of protozoan in ecotoxicology. **Environmental Pollution**, v. 213, p. 160–172, 2016.

MELO, S. M.; BRITO, N. M. Analysis and Occurrence of Endocrine Disruptors in Brazilian Water by HPLC-Fluorescence Detection. **Water, Air, & Soil Pollution**, v. 225, n. 1, p. 1783, 6 jan. 2014.

MUNZ, N. A. et al. Pesticides drive risk of micropollutants in wastewater-impacted streams during low flow conditions. **Water Research**, v. 110, p. 366–377, 2017.

NEALE, P. A. et al. Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. **Water Research**, v. 123, p. 734–750, 2017.

NICOLAS, C. I. et al. Rapid experimental measurements of physicochemical properties to inform models and testing. **Science of the Total Environment**, v. 636, p. 901–909, 2018.

NÖDLER, K.; VOUTSA, D.; LICHA, T. Polar organic micropollutants in the coastal environment of different marine systems. **Marine Pollution Bulletin**, v. 85, n. 1, p. 50–59, 2014.

NORTON, S. B. et al. A framework for ecological risk assessment at the EPA.

- Environmental Toxicology and Chemistry**, v. 11, n. 12, p. 1663–1672, 1992.
- OECD. Disponível em: <<https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm>>.
- PETRIE, B.; BARDEN, R.; KASPRZYK-HORDERN, B. A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. **Water Research**, v. 72, n. 0, p. 3–27, 2015.
- PORCELLI, C. et al. A comparison of DEMETRA individual QSARs with an index for evaluation of uncertainty. **Chemosphere**, v. 71, n. 10, p. 1845–1852, 2008.
- RAPPAPORT, S. M. Implications of the exposome for exposure science. **Journal of Exposure Science & Environmental Epidemiology**, v. 21, n. 1, p. 5–9, 17 jan. 2011.
- RIVERA-UTRILLA, J. et al. Pharmaceuticals as emerging contaminants and their removal from water. A review. **Chemosphere**, v. 93, n. 7, p. 1268–1287, 2013.
- SANTHOSH, C. et al. Role of nanomaterials in water treatment applications: A review. **Chemical Engineering Journal**, v. 306, p. 1116–1137, 2016.
- SIPPL, W. **3D-QSAR – Applications, recent advances, and limitations**. [s.l.: s.n.]. v. 8
- TAVAKOL, M. et al. Determining multivariate analysis sampling frequency for monitoring contamination caused by trout farms. **Polish Journal of Environmental Studies**, v. 26, n. 1, p. 337–346, 2017.
- TODESCHINI, R. **Molecular Descriptors for Chemoinformatics, Volumes I & II**. [s.l.: s.n.].
- TOOLARAM, A. P. et al. Initial hazard screening for genotoxicity of photo-transformation products of ciprofloxacin by applying a combination of experimental and in-silico testing. **Environmental Pollution**, v. 211, p. 148–156, 2016.
- TOXICITY, E. User ' s Guide for T . E . S . T . (Toxicity Estimation Software Tool) Table of Contents. **Program**, 2008.
- WANG, Q.; WANG, P.; YANG, Q. Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. **Science of the Total Environment**, v. 621, p. 990–999, 2018.
- YANG, Y. et al. Occurrences and removal of pharmaceuticals and personal care products (PPCPs) in drinking water and water/sewage treatment plants: A review. **Science of the Total Environment**, v. 596–597, p. 303–320, 2017.

CAPÍTULO II

CAPÍTULO II - PREDIÇÃO DE ECOTOXICIDADE DE QUÍMICOS ORGÂNICOS CLASSIFICADOS COMO MOA 1 USANDO ABORDAGEM *IN SILICO*¹

RESUMO

Os produtos orgânicos persistentes são compostos utilizados para diversos fins, como produtos de higiene pessoal, surfactantes, corantes, aditivos industriais, alimentos, pesticidas e produtos farmacêuticos. Essas substâncias são constantemente introduzidas no meio ambiente e muitos desses poluentes são difíceis de degradar. Métodos *in silico*, como estudos de Relação Quantitativa entre Estrutura Química e Atividade (QSARs), têm sido utilizados para desenvolver modelos importantes de predição em várias áreas da ciência, bem como estudos de toxicidade aquática. O objetivo do presente estudo foi construir um conjunto de descritores moleculares teóricos baseados em QSAR dos compostos definidos como MoA 1 (Modo de Ação 1), produtos químicos não reativos quando considerando efeitos agudos e que não interagem com receptores específicos em um organismo, para identificar a relação entre as propriedades moleculares e os valores de toxicidade aguda em peixes. Os resultados foram selecionados a partir do modelo PLS (*partial least squares*), baseados nos valores dos coeficientes de determinação por validação cruzada, $Q_{cv}^2 = 0,793$, coeficiente de determinação (R^2) = 0,823, variância explicada na predição externa (Q_{ext}^2) = 0,87. Dos descritores selecionados, não só a hidrofobicidade está relacionada à toxicidade, mas outras propriedades físico-químicas combinadas contribuem para a atividade desses compostos. Como a distribuição simétrica das porções hidrofóbicas na estrutura dos compostos, bem como a forma, como cadeias ramificadas são características importantes relacionadas à toxicidade. Por outro lado, os compostos menos tóxicos apresentam superfícies hidrofóbicas assimetricamente distribuídas em torno das moléculas. Neste estudo, uma tendencia estrutural para toxicidade foi destacada pelo modelo PLS, permitindo a classificação de compostos MoA 1 mais tóxicos.

Palavras-chave: QSAR; Toxicidade aguda dos peixes; Descritores Moleculares; Software Volsurf; MoA 1.

¹ Este capítulo aparece previamente como artigo na revista *Ecotoxicology and Environmental Safety* 153 (2018) 151-159.

ABSTRACT

Persistent organic products are compounds used for various purposes, such as personal care products, surfactants, dyes, industrial additives, food, pesticides and pharmaceutical products. These substances are constantly introduced into the environment and many of these pollutants are difficult to degrade. In silico methods, such as studies of Quantitative Relationship between Chemical Structure and Activity (QSARs), have been used to develop important models of prediction in several areas of science, as well as studies of aquatic toxicity. The objective of the present study was to build a set of theoretical molecular descriptors based on QSAR of the compounds defined as MoA 1 (Mode of Action 1), non-reactive chemicals when considering acute effects and that do not interact with specific receptors in an organism, to identify the relationship between molecular properties and acute toxicity values in fish. The results were selected from the PLS model (partial least squares), based on the values of the coefficients of determination by cross-validation, $Q_{cv2} = 0.793$, coefficient of determination (R^2) = 0.823, variance explained in the external prediction (Q_{ext2}) = 0, 87. Of the selected descriptors, not only is hydrophobicity related to toxicity, but other combined physicochemical properties contribute to the activity of these compounds. As the symmetrical distribution of the hydrophobic portions in the structure of the compounds, as well as the shape, as branched chains are important characteristics related to toxicity. On the other hand, less toxic compounds have hydrophobic surfaces asymmetrically distributed around the molecules. In this study, a structural tendency towards toxicity was highlighted by the PLS model, allowing the classification of more toxic MoA 1 compounds.

Keywords: QSAR; Acute toxicity of fish; Molecular descriptors; Volsurf software; MoA 1.

1 INTRODUÇÃO

Os compostos orgânicos representam uma ampla gama de produtos químicos com diferentes propriedades físico-químicas, seja de fontes sintéticas ou naturais. Estes incluem produtos orgânicos persistentes, que são compostos utilizados para vários fins, como produtos de higiene pessoal, surfactantes, corantes, aditivos industriais, alimentos, pesticidas e produtos farmacêuticos (ISIDORI *et al.*, 2016; RIBEIRO *et al.*, 2015). Essas substâncias são constantemente introduzidas no meio ambiente, originadas principalmente dos lançamentos de fontes industriais, domésticas, hospitalares e de efluentes de superfície das áreas agrícolas e pecuárias (ESTEBAN *et al.*, 2014; MOUSE *et al.*, 2017; SERRA-ROIG *et al.*, 2016). A sua ocorrência é amplamente detectada em diferentes matrizes aquáticas, águas superficiais, águas subterrâneas e oceanos (BUSCH *et al.*, 2016; GROS *et al.*, 2016, 2012). Muitos desses poluentes são difíceis de degradar, devido à sua alta estabilidade química.

A exposição de contaminantes químicos ao meio aquático representa sérias ameaças à preservação da qualidade ambiental e é reconhecida por especialistas de vários países como um grande problema global. Consequentemente, eles demonstram efeitos sobre a manutenção dos ecossistemas aquáticos, bem como sobre questões relacionadas à saúde humana (BOURGIN *et al.*, 2013). Efeitos adversos já foram identificados e continuam relacionados à presença de poluentes na água, como a inibição de neurotransmissores (de Oliveira *et al.*, 2016), efeitos mutagênicos (LUTTERBECK *et al.*, 2015), carcinogênico, biomagnificação em redes trófica marinha (XUE *et al.*, 2017), fitotoxicidade (RICHTER *et al.*, 2016), feminização de peixes (HICKS *et al.*, 2017) e desenvolvimento de bactérias resistentes a antibióticos (MIRANDA *et al.*, 2016).

Para permitir a determinação do mecanismo pelo qual uma substância causa toxicidade, os estudos desenvolvidos em toxicologia utilizaram modelos de similaridade estrutural ou grupos funcionais com o objetivo de classificar produtos químicos para seu modo de ação (MoA) (NENDZA *et al.*, 2014). O MoA pode ser descrito como uma característica estrutural que pode ser responsável pelas propriedades tóxicas de um composto. Em outras palavras, a maneira pela qual um tóxico afeta de forma funcional ou anatômica um organismo vivo (IORIO *et al.*, 2010; SINGH *et al.*, 2016).

Quatro classes de MoA são categorizadas de acordo com VERHAAR *et al.*, 1992: MoA 1 - corresponde a compostos químicos com mecanismo de ação tóxico mais básico, que atuam como narcose não polar, não reativos quando considerando efeitos agudos e que não interage com receptores específicos nos organismos; MoA 2 - atuam como narcose polar, que apresentam baixa toxicidade e geralmente são caracterizados por possuírem acidez doadora de ligação de hidrogênio; MoA 3 - são produtos químicos reativos que reagem de forma não seletiva e MoA 4 - substâncias especificamente ativas que exibem toxicidade devido a interações específicas com certas moléculas receptoras. Os compostos que não podem ser classificados como pertencentes às classes 1, 2 ou 3 e que não são conhecidos por agir por um mecanismo específico só podem ser classificados como "não é possível classificar de acordo com essas regras" (VERHAAR *et al.*, 1992).

Métodos *in silico*, tais como [(Q)SARs] [do inglês (*Quantitative*) *Structure-Activity Relationships*] têm sido utilizados para desenvolver modelos importantes para a predição em várias áreas da ciência, bem como estudos de toxicidade aquática (CASSOTTI *et al.*, 2015). As técnicas QSAR fornecem um meio alternativo importante e útil de teste; eles podem ser empregados para lidar com grandes quantidades de dados, para avaliar os perigos potenciais de produtos químicos, evitar experiências caras e demoradas, bem como alternativas para limitar testes experimentais em animais (CHEN *et al.*, 2015; GRAMÁTICA *et al.*, 2016). Os modelos computacionais têm sido introduzidos na avaliação da segurança química e no suporte à decisão regulamentar exigido pela legislação da União Europeia no Registro, Avaliação, Autorização e restrição de Químicos (REACH) e no Estados Unidos pela Agência de Proteção Ambiental (US EPA), para proteção da saúde pública e do meio ambiente (DELGADO *et al.*, 2012, ECHA, 2014, US EPA, 2016).

O objetivo do presente estudo foi desenvolver modelo baseado em QSAR usando um conjunto de descritores moleculares gerados a partir dos valores de toxicidade aguda para peixes de compostos definidos como MoA 1. A fim de identificar as propriedades moleculares relacionadas a este mecanismo e prever a toxicidade para peixes de compostos não testados. O modelo QSAR foi validado de acordo com os princípios de validação, para fins regulatórios e a aceitabilidade de (Q)SARs, proposto pela Organização de Cooperação e Desenvolvimento Econômico - OCDE (OECD, 2007).

2 MATERIAL E MÉTODOS

2.1 Conjunto de dados

Para o desenvolvimento do modelo QSAR, o conjunto de dados compreendeu 61 compostos orgânicos, obtidos a partir do banco de dados relatado na literatura contendo dados toxicológicos experimentais de compostos químicos classificados como MoA1 (narcose não polar) (THOMAS *et al.*, 2015). Os dados experimentais utilizados no presente estudo foi formado por 36 compostos químicos e 25 compostos foram empregados para validação externa com base nas previsões realizadas para o conjunto de testes independente com o objetivo de avaliar a robustez dos modelos desenvolvidos (KONEMANN, 1981; KONEMANN and MUSCH, 1981). Os valores experimentais determinados por EC₅₀ (concentração efetiva) (mol/L) foram convertidos em uma base molar e, em seguida, os dados logaritmicamente transformados [$-\log EC_{50} = pEC_{50}$] foram utilizados como variáveis de resposta.

2.2 Descritores moleculares

Para gerar os descritores moleculares, os SMILES (do inglês, *simplified molecular-input line-entry system*) das estruturas químicas foram convertidos em um arquivo .mol (formato MDL – *Molecular Design Ltd*). A ferramenta Standardizer foi utilizada para converter as estruturas químicas em representações canônicas, [JChem 16.1.11.0, 2016], ChemAxon (<http://www.chemaxon.com>), garantindo a comparabilidade de todas as representações moleculares. As padronizações aplicadas a cada estrutura foram extração de sais, adição de hidrogênios explícitos e aromatização das moléculas. E também foi usada para gerar as estruturas tridimensionais (3D).

Através do programa VolSurf + v. 1.0.7 (<http://www.moldiscovery.com>), as estruturas 3D dos compostos foram usadas como dados de entrada e para cálculo de campos de interação molecular (MIF – molecular interaction field) para gerar descritores moleculares usando as seguintes sondas: N1 (nitrogênio amida), O (oxigênio de carbonila), H₂O (sonda de água) e DRY (sonda hidrofóbica) (CRUCIANI *et al.*, 2000). Descritores não-derivados de

MIF foram calculados, incluindo descritores que quantificam o tamanho molecular, a forma, as regiões hidrofílicas e hidrofóbicas, momentos de energia de interação, fatores de capacidade, equilíbrio hidrofóbico-lipofílico e outros descritores. Estes descritores foram selecionados para o presente estudo pois são simples de usar, de fácil interpretação e compreensão do mecanismo de ação e/ou do significado físico.

2.3 Análise estatística e validação do modelo

Os modelos QSAR foram obtidos usando PLS (do inglês, *Partial Least Squares*) para explorar e extrair informações (BARONI *et al.*, 1993; WOLD *et al.*, 1984). O pré-processamento automático para todas as variáveis independentes foi aplicado. O número de variáveis originais foi selecionado usando um gráfico de VIP (do inglês, *Variable Influence on Projection*). O gráfico VIP mostra a importância das variáveis independentes, quer para as variâncias explicadas entre si, quer para a variável dependente no modelo PLS, enquanto os coeficientes PLS indicam a contribuição de cada descritor para o modelo (ERMONDI *et al.*, 2011).

Os modelos foram avaliados pelo coeficiente de determinação (R^2) e o coeficiente de determinação por validação cruzada “leave-one-out” (Q_{cv}^2). O número de Variáveis Latentes (VLs) e o modelo foram selecionados usando o maior valor do Q_{cv}^2 .

2.4 Domínio de Aplicabilidade (DA)

O domínio de aplicabilidade (DA), é calculado para testar as limitações de um modelo, ou seja, a faixa de propriedades ou estruturas moleculares para as quais o modelo é considerado aplicável (TROP SHA, 2010). Neste estudo foi calculado com base nas distâncias Euclidianas, a fim de definir um domínio químico baseado na gama de propriedades moleculares ou estruturas das moléculas de treinamento para as quais o modelo é considerado aplicável. Os compostos no conjunto de testes para os quais as previsões forneçam valores superiores aos valores médios de distância euclidiana de todas as amostras no conjunto de treinamento podem não ser confiáveis (Equação 1). O parâmetro Z é um valor

de corte empírico, definido pelo usuário, 0.5 foi usado como padrão (AFANTITIS *et al.*, 2011; MELAGRAKI *et al.*, 2010).

$$DA = d + Z\sigma$$

Equação 1

onde: d e σ são a distância euclidiana média e o desvio padrão do conjunto de treinamento.

3 RESULTADOS E DISCUSSÕES

3.1 Geração de Modelo

Os dados previstos e experimentais de toxicidade para peixe, seus valores residuais para os modelos de treinamento e conjuntos de teste são apresentados na Tabela 1 e Tabela 2. O banco de dados de toxicidade aguda em peixes foi composto de um conjunto de treinamento com 36 compostos químicos (59,01%) com uma resposta experimental variando entre 2,12 e 5,76 e um conjunto de teste com 25 compostos químicos (40,98%).

Tabela 1 - Estruturas químicas no formato SMILES e valores experimentais de pEC₅₀ para predição externa e respectivos erros

Id ^a	SMILES	pEC ₅₀ experimental	pEC ₅₀ predito externo	Erro (predito- experimental)
87	<chem>CC(Cl)(Cl)Cl</chem>	3	3.56	0.56
88	<chem>CC(Cl)Cl</chem>	2.69	2.90	0.21
89	<chem>Clc1ccc(Cl)c(Cl)c1Cl</chem>	5.43	5.32	-0.11
90	<chem>Clc1cc(Cl)c(Cl)c(Cl)c1</chem>	5.43	5.31	-0.12
91	<chem>Clc1cccc(Cl)c1Cl</chem>	4.89	4.85	-0.04
92	<chem>ClCC(Cl)CCl</chem>	3.55	4.01	0.46
93	<chem>Clc1cccc1Cl</chem>	4.4	4.44	0.04
94	<chem>C[C@H](Cl)CCl</chem>	2.99	3.41	0.42
95	<chem>Clc1cc(Cl)cc(Cl)c1</chem>	4.74	4.83	0.09
96	<chem>Clc1cccc(Cl)c1</chem>	4.3	4.41	0.11

97	<chem>C1CCCCI</chem>	3.13	3.62	0.49
98	<chem>Clc1ccc(Cl)cc1</chem>	4.57	4.38	-0.19
99	<chem>C1C=C\CCI</chem>	3.5	3.99	0.49
100	<chem>C1CCCCCCI</chem>	4.1	4.34	0.24
101	<chem>CCCCCC1</chem>	2.98	3.04	0.06
102	<chem>C1CCOCCCC1</chem>	3.42	3.96	0.54
103	<chem>C1CC(Cl)=C</chem>	4	3.61	-0.39
104	<chem>Cc1cc(Cl)c(Cl)cc1Cl</chem>	5.06	5.09	0.03
105	<chem>Nc1ccc(Cl)cc1Cl</chem>	4.14	4.19	0.05
106	<chem>Cc1ccc(Cl)cc1Cl</chem>	4.54	4.64	0.10
107	<chem>Cc1ccc(Cl)c(Cl)c1</chem>	4.5	4.62	0.12

108	<chem>Cc1cccc(Cl)c1</chem>	3.84	4.13	0.29
109	<chem>Cc1ccc(Cl)cc1</chem>	4.33	4.12	-0.21
110	<chem>Clc1ccccc1</chem>	3.77	3.93	0.16
111	<chem>ClC(Cl)C(Cl)(Cl)Cl</chem>	4.13	4.62	0.49

^a Identificação.

Tabela 2 - Estruturas químicas no formato SMILES usada no conjunto de treino aplicado no modelo PLS neste estudo

Id ^a	SMILES	p EC ₅₀ experimental	pEC ₅₀ predicted	Errors (predicted - experimental)
1	<chem>CCCCCO</chem>	3.02	3.09	0.06
2	<chem>CCCCCCO</chem>	3.49	3.57	0.08
3	<chem>CCCCCCC</chem>	4.00	3.77	-0.23
4	<chem>CCCCCCCC</chem>	4.42	4.28	-0.14
5	<chem>CCCCCCCCC</chem>	4.84	4.53	-0.30
6	<chem>CCCCCCCCC</chem>	5.24	4.88	-0.36
7	<chem>CCCCCCCCC</chem>	5.27	5.13	-0.14
8	<chem>CC(C)CCCCCCCC</chem>	5.56	5.46	-0.11
9	<chem>OC1CCCCC1</chem>	2.15	2.98	0.83

10	<chem>OCc1ccccc1</chem>	2.37	3.06	0.69
11	<chem>COC(C)(C)C</chem>	2.12	2.24	0.12
12	<chem>COC(C)(C)C</chem>	2.19	2.24	0.05
13	<chem>CICCl</chem>	2.64	2.74	0.10
14	<chem>CICCl</chem>	2.93	2.74	-0.19
15	<chem>CICCl</chem>	2.41	2.74	0.33
16	<chem>CIC(Cl)Cl</chem>	3.82	3.33	-0.48
17	<chem>CIC(Cl)Cl</chem>	3.06	3.33	0.27
18	<chem>CIC(Cl)Cl</chem>	3.37	3.33	-0.04
19	<chem>CIC(Cl)Cl</chem>	3.20	3.33	0.13
20	<chem>CIC(Cl)Cl</chem>	3.23	3.33	0.11
21	<chem>CIC(Cl)Cl</chem>	2.99	3.33	0.34
22	<chem>CIC(Cl)(Cl)Cl</chem>	3.80	3.81	0.01
23	<chem>CIC=C(Cl)Cl</chem>	3.92	3.79	-0.13
24	<chem>CIC=C(Cl)Cl</chem>	3.67	3.79	0.12
25	<chem>CIC=C(Cl)Cl</chem>	3.51	3.79	0.28
26	<chem>CIC=C(Cl)Cl</chem>	3.47	3.79	0.31
27	<chem>CIC(Cl)=C(Cl)Cl</chem>	4.52	4.29	-0.23
28	<chem>CIC(Cl)=C(Cl)Cl</chem>	4.52	4.29	-0.23
29	<chem>CIC(Cl)=C(Cl)Cl</chem>	4.09	4.29	0.20
30	<chem>CIC(Cl)=C(Cl)Cl</chem>	3.84	4.29	0.45

31	<chem>C1C(Cl)=C(Cl)Cl</chem>	4.30	4.29	0.00
32	<chem>C1CCCCl</chem>	2.86	2.89	0.03
33	<chem>C1CCCCl</chem>	2.94	2.89	-0.04
34	<chem>C1CCCCl</chem>	2.92	2.89	-0.03
35	<chem>Clc1ccc(Cl)c(Cl)c1</chem>	4.88	4.84	-0.04
36	<chem>C1C1=CC(Cl)=CC=C1</chem>	4.28	4.51	0.23
37	<chem>C1C1=CC(Cl)=CC=C1</chem>	4.47	4.51	0.04
38	<chem>C1C1=CC(Cl)=CC=C1</chem>	4.21	4.51	0.30
39	<chem>C1C1=CC=C(Cl)C=C1</chem>	5.12	4.59	-0.53
40	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.54	4.59	0.05
41	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.10	4.59	0.49
42	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.54	4.59	0.05
43	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.42	4.59	0.17
44	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.85	4.59	-0.26
45	<chem>C1C1=CC=C(Cl)C=C1</chem>	4.85	4.59	-0.26
46	<chem>Clc1cccc1Cl</chem>	4.99	4.48	-0.50
47	<chem>Clc1cccc1Cl</chem>	4.97	4.48	-0.48
48	<chem>Clc1cccc1Cl</chem>	4.96	4.48	-0.48
49	<chem>Clc1cccc1Cl</chem>	4.98	4.48	-0.49
50	<chem>Clc1cccc1Cl</chem>	4.45	4.48	0.03
51	<chem>C1C(Cl)C(Cl)Cl</chem>	3.92	4.25	0.33

52	<chem>C1C(Cl)C(Cl)Cl</chem>	3.92	4.25	0.33
53	<chem>C1C(Cl)C(Cl)Cl</chem>	3.96	4.25	0.29
54	<chem>C1C(Cl)C(Cl)Cl</chem>	3.80	4.25	0.45
55	<chem>C1C1=CC=CC(Cl)=C1Cl</chem>	5.72	4.88	-0.83
56	<chem>C1C1=CC=CC(Cl)=C1Cl</chem>	4.75	4.88	0.13
57	<chem>CC(Cl)(Cl)Cl</chem>	3.40	3.52	0.11
58	<chem>CC(Cl)(Cl)Cl</chem>	4.08	3.52	-0.56
59	<chem>CC(Cl)(Cl)Cl</chem>	3.45	3.52	0.07
60	<chem>CC(Cl)(Cl)Cl</chem>	3.51	3.52	0.00
61	<chem>CC(Cl)(Cl)Cl</chem>	3.38	3.52	0.14
62	<chem>CC(Cl)(Cl)Cl</chem>	3.61	3.52	-0.09
63	<chem>CC(Cl)(Cl)Cl</chem>	3.50	3.52	0.02
64	<chem>CC(Cl)(Cl)Cl</chem>	3.67	3.52	-0.15
65	<chem>C1CC(Cl)Cl</chem>	3.52	3.97	0.45
66	<chem>Clc1ccccc1</chem>	4.40	3.96	-0.44
67	<chem>CCCCCC</chem>	4.23	3.32	-0.91
68	<chem>C1CCCCC1</chem>	4.27	3.29	-0.98
69	<chem>CCCCC=C</chem>	4.18	3.37	-0.81
70	<chem>COC(=O)c1ccccc1C(=O)OC</chem>	3.59	3.87	0.28
	<chem>CCOC(=O)c1ccccc1C(=O)O</chem>			
71	<chem>CC</chem>	4.12	4.47	0.34

	<chem>CCOC(=O)c1ccccc1C(=O)O</chem>			
72	<chem>CC</chem>	4.12	4.47	0.35
	<chem>CCOC(=O)c1ccccc1C(=O)O</chem>			
73	<chem>CC</chem>	4.27	4.47	0.20
	<chem>CCOC(=O)c1ccccc1C(=O)O</chem>			
74	<chem>CC</chem>	3.88	4.47	0.58
	<chem>CCOC(=O)c1ccccc1C(=O)O</chem>			
75	<chem>CC</chem>	4.12	4.47	0.34
	<chem>CCCCOC(=O)c1ccccc1C(=O)O</chem>			
76	<chem>O)OCCCC</chem>	5.26	5.63	0.37
	<chem>CCCCOC(=O)c1ccccc1C(=O)O</chem>			
77	<chem>O)OCCCC</chem>	5.48	5.63	0.15
	<chem>CCCCOC(=O)c1ccccc1C(=O)O</chem>			
78	<chem>O)OCCCC</chem>	5.24	5.63	0.39
	<chem>CCCCOC(=O)c1ccccc1C(=O)O</chem>			
79	<chem>O)OCCCC</chem>	5.76	5.63	-0.14
80	<chem>O=N(=O)c1ccccc1</chem>	3.13	3.26	0.14
81	<chem>O=N(=O)c1ccccc1</chem>	3.32	3.26	-0.06
82	<chem>O=N(=O)c1ccccc1</chem>	3.02	3.26	0.25
83	<chem>Cc1ccccc1N(=O)=O</chem>	3.66	3.48	-0.18
84	<chem>Cc1cccc(c1)N(=O)=O</chem>	3.63	3.56	-0.07
85	<chem>Cc1cccc(c1)N(=O)=O</chem>	4.27	3.56	-0.71
86	<chem>Cc1ccc(cc1)N(=O)=O</chem>	3.54	3.55	0.02

^a Identificação.

O modelo PLS selecionado para toxicidade aguda de peixes apresentou os seguintes valores de coeficientes estatísticos: $Q_{cv}^2 = 0,793$, Equação 1, desvio padrão dos erros da predição (SDEP) = 0,381, Equação 3, $R^2 = 0,823$ (Figura 1), Equação 1 e desvio padrão dos erros do calculado (SDEC) = 0,352, Equação 4. A validação externa composta por 25 compostos (Tabela 1) resultou nos seguintes valores: variação explicada na previsão (Q_{ext}^2) = 0,87 e desvio padrão de erros de previsão externa ($SDEP_{ext}$) = 0,29. A equação PLS selecionada tem apenas dois VLs, que compreendem 22 descritores moleculares (S3), porque VLs no modelo não proporcionou uma melhoria significativa no valor Q_{cv}^2 .

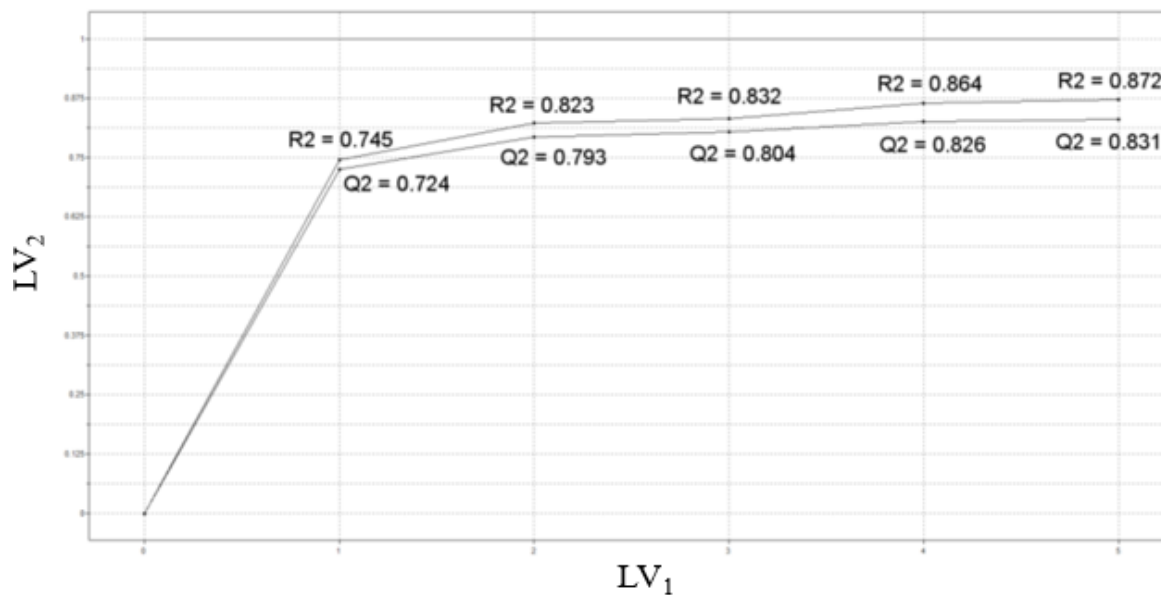
$$R^2 / Q_{cv}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} \quad \text{Equação 1}$$

$$SDEP = \sqrt{\frac{PRESS}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_{i,-i})^2} \quad \text{Equação 2}$$

$$SDEC = \sqrt{\frac{RSS}{n_{tr}}} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}} \quad \text{Equação 3}$$

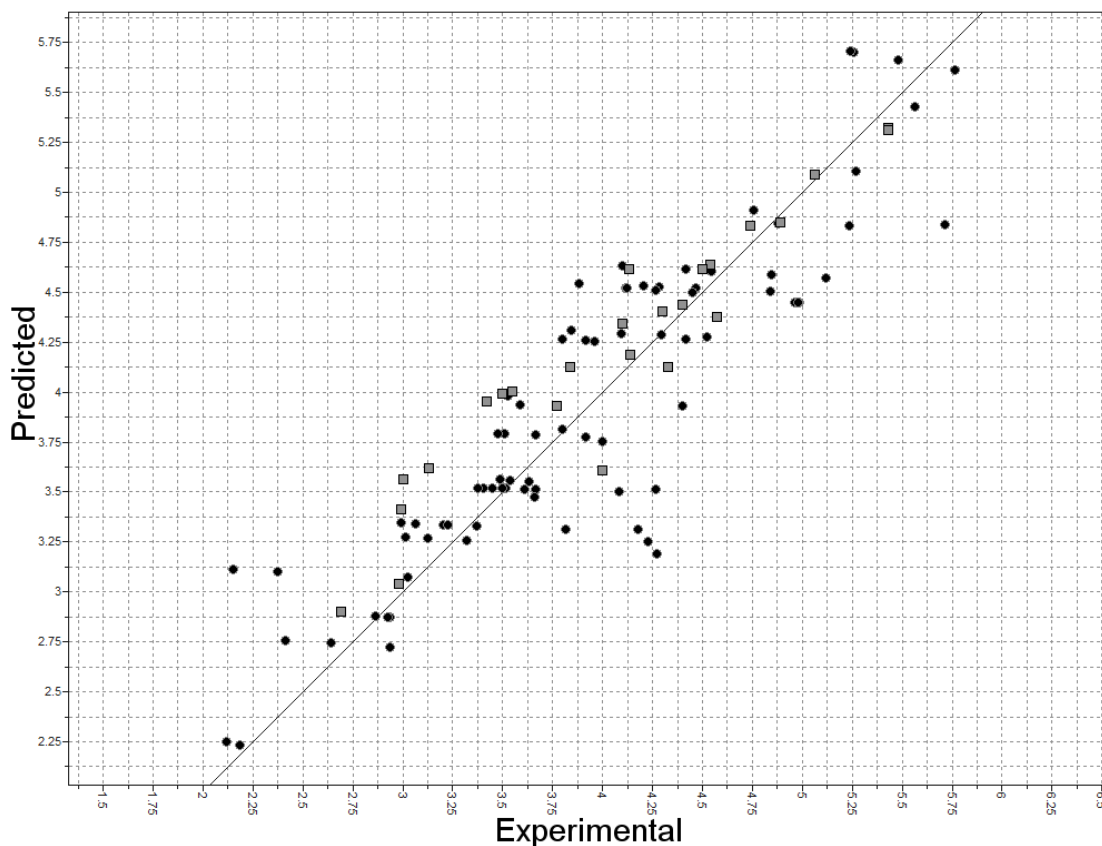
onde y_i é a resposta experimental da i -ésima molécula e \bar{y} é a resposta média dos compostos químicos usados no treinamento. \hat{y}_i é a resposta calculada quando a i -ésima molécula foi incluída no conjunto de treinamento para o caso de ajuste (R^2) ou prevista quando a i -ésima molécula fazia parte do conjunto de avaliação para validação cruzada (Q_{cv}^2).

Figura 1 - Gráfico de R^2 e Q^2 versus o número de VLs do modelo PLS. O modelo foi selecionado contendo duas variáveis ($R^2 = 0,823$ and $Q_{cv}^2 = 0,793$).



A validação externa pode ser aplicada de forma útil pelo modelo QSAR para definir com mais precisão o seu poder preditivo real (GRAMATICA, 2007). Para validação externa, foram utilizados 25 compostos, presentes nas Tabela 1. Na Figura 2, o gráfico de valores experimentais vs. preditos, para validação cruzada e conjunto de teste externo, descrevem a distribuição dos pontos em relação às linhas de melhor ajuste e principalmente diferencia os compostos mais tóxicos. O domínio de aplicabilidade com base nas distâncias euclidianas foi utilizado para verificar se as previsões dos compostos no conjunto de treinamento do modelo desenvolvido são confiáveis.

Figura 2 - Gráfico PLS dos valores predito vs. experimentais para os dados de toxicidade aguda dos peixes. Os compostos externos são representados por quadrados cinzas e o conjunto de treinamento são os pontos em negrito.



3.2 Modelo de refinamento e interpretação química

O gráfico de escores para o modelo resultante com os dois VLs (VL1 versus VL2), para os conjuntos de treino e teste, são apresentados nas Figuras 3 e 4. Ambos os gráficos proporcionam uma boa discriminação entre os tipos de compostos mais tóxicos (pontos em negrito) e menos tóxicos (pontos mais claros). Os pontos em negrito no primeiro quadrante do gráfico representam o espaço químico onde os compostos possuem valores mais altos de pEC_{50} , enquanto os pontos brancos à esquerda, em maior parte no terceiro quadrante, representam compostos com menor pEC_{50} .

Figura 3 - Gráfico de escores dos 37 compostos do conjunto de treino (VL1 versus VL2) do modelo PLS. Os pontos em negrito representam os compostos com maiores valores de toxicidade aguda de peixes.

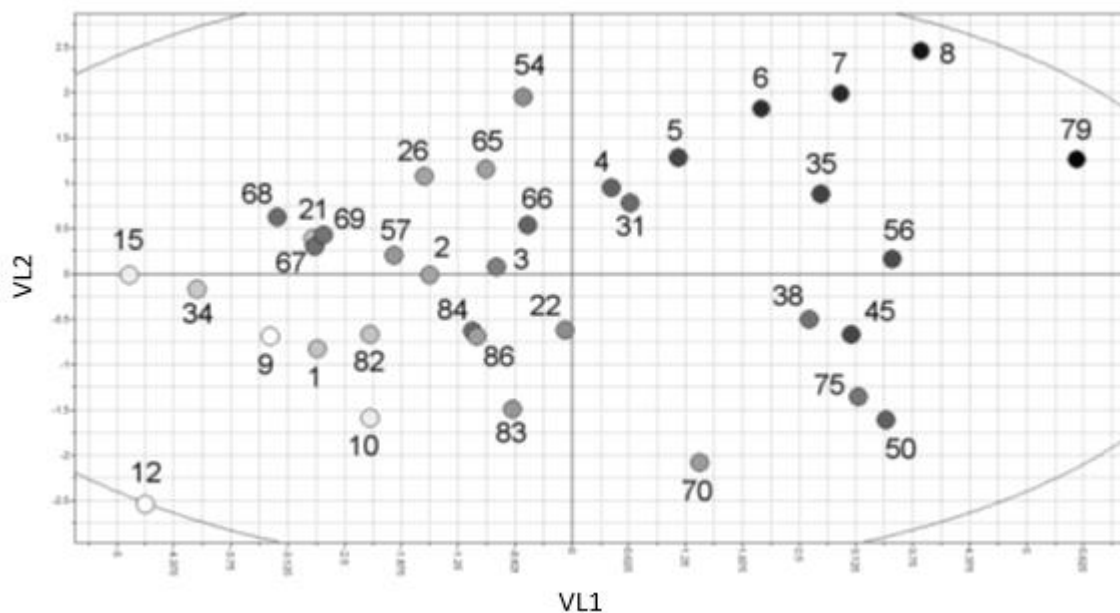
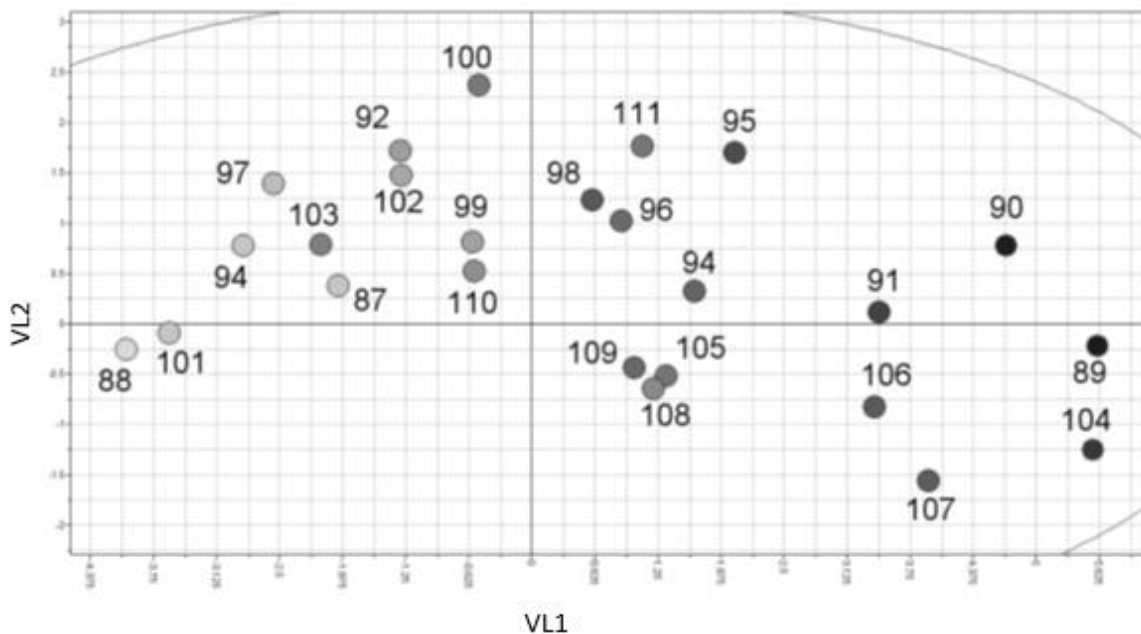


Figura 4 - Gráfico de escores com 25 compostos do conjunto de validação externa (VL1 versus VL2) do modelo PLS. Os pontos mais escuros representam os compostos com maiores valores de toxicidade aguda de peixes.



3.3 Interpretação do modelo

O gráfico de pesos de VL1 versus VL2 (Figura 5), indica os descritores moleculares relevantes no modelo para todos os compostos químicos. A região dos descritores no gráfico está relacionada aos compostos químicos colocados na mesma região. Os descritores distribuídos no primeiro quadrante têm maior contribuição para os valores de pEC₅₀. A maioria das moléculas mostram contribuições positiva para o valor de pEC₅₀. Valores mais elevados dos coeficientes dos descritores representam uma maior contribuição para a atividade, Tabela 3.

Os descritores moleculares mais relevantes no modelo, com contribuição positiva para o primeiro e segundo componentes do PLS (VL1 e VL2) e conseqüentemente para os valores de pEC₅₀, são D1 (coeficiente 0,216), D2 (coeficiente 0,168), área de superfície hidrofóbica (ASH) (coeficiente 0.136) e R (coeficiente 0.120). Os descritores D1 e D2 são gerados pela sonda hidrofóbica e são definidos como o envelope molecular gerando interações hidrofóbicas atraentes, calculadas a níveis de energia de -0.2 e -0.4 kcal / mol respectivamente. Portanto, os compostos com volumes hidrofóbicos com esses níveis de energia são mais tóxicos. Outro descritor com uma significativa contribuição positiva para valores de pEC₅₀ que confirma a toxicidade de compostos mais hidrofóbicos é a ASH, que se refere à soma das contribuições da região hidrofóbica.

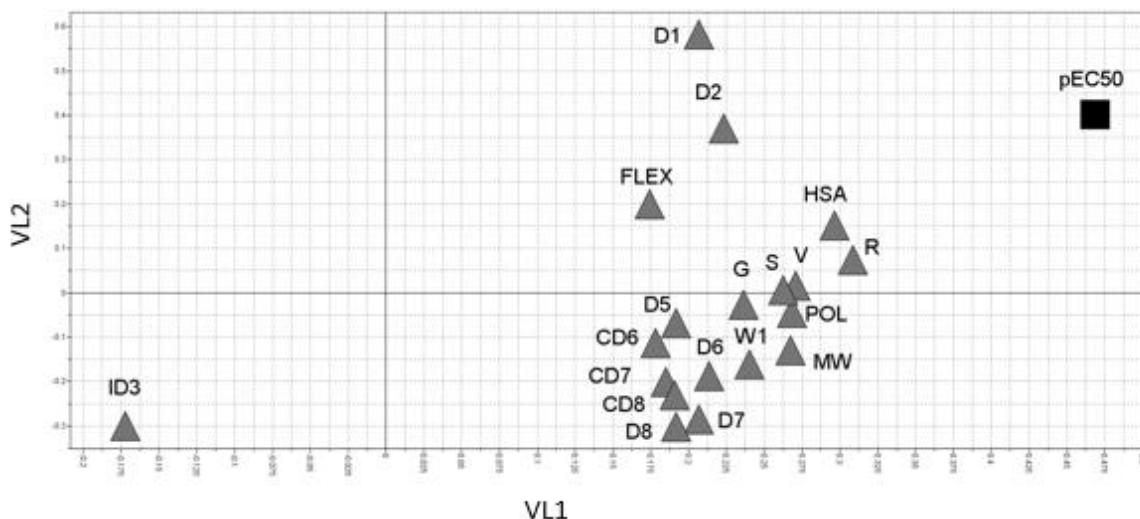
O descritor R depende da medida de rugosidade da superfície molecular e é representado pela proporção de volume/superfície da molécula. Quanto menor a proporção, maior a rugosidade, assim, compostos ramificados com mais substituintes têm valores mais altos de R; em contrapartida, compostos não-substituídos, linear ou cíclico apresentam menor rugosidade.

O único descritor com contribuição negativa significativa no modelo foi o ID3 (coeficiente -0.133). ID3 expressa o desequilíbrio entre o centro de massa de uma molécula e o baricentro de suas regiões hidrofóbicas. A contribuição negativa deste descritor nos informa que os compostos menos tóxicos têm o centro de massa próximo ao baricentro das regiões hidrofóbicas.

Tabela 3 - Descritores moleculares selecionados no modelo QSAR

	Descritores	Definição	Valores de coeficientes
1	D1	Hydrophobic regions	0.216
2	D2	Hydrophobic regions	0.168
3	HSA	Polar and Hydrophobic Surface Areas	0.136
4	R	Volume/surface ratio	0.120
5	FLEX	Flexibility parameters	0.108
6	V	Molecular Volume	0.093
7	S	Molecular Surface area	0.088
8	POL	Polarizability	0.077
9	G	Molecular Globularity	0.071
10	MW	Molecular Weight	0.055
11	D5	Hydrophobic regions	0.046
12	W1	Hydrophilic regions	0.038
13	CD6	Capacity Factor for DRY	0.030
14	D6	Hydrophobic regions	0.022
15	CD7	Capacity Factor for DRY	0.010
16	CD8	Capacity Factor for DRY	0.004
17	WO5	H-bond donor volumes	0.000
18	WO6	H-bond donor volumes	0.000
19	NCC	Number of Charged Centers	0.000
20	D7	Hydrophobic regions	-0.005
21	D8	Hydrophobic regions	-0.014
22	ID3	Hydrophobic Integy moment	-0.133

Figura 5 - Gráfico de peso (VL1 versus VL2) do modelo PLS. Os rótulos dos descritores moleculares (triângulos) referem-se aos símbolos listados na Tabela 3. O quadrado representa a variável dependente, pEC₅₀, o logaritmo negativo da EC₅₀ - 50% da concentração efetiva (mol/L) para a toxicidade aguda dos peixes.



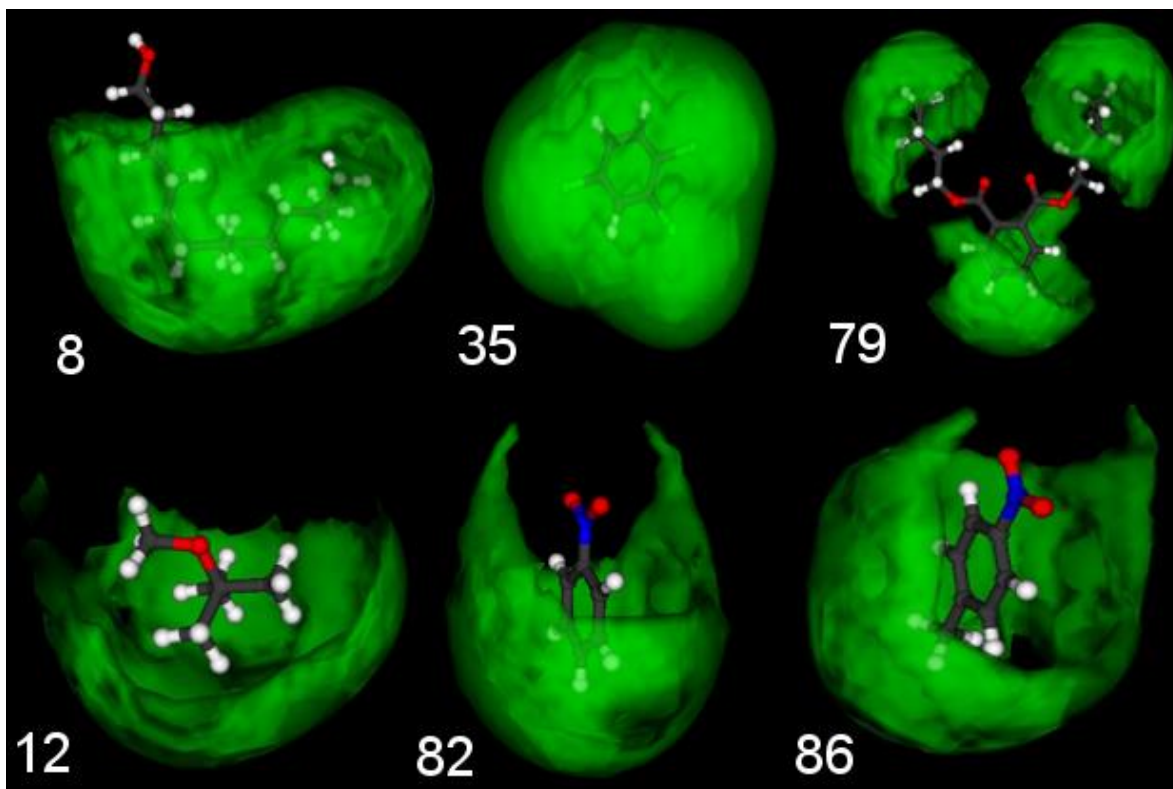
Os compostos 8 (isotridecanol), 35 (1,2,4-triclorobenzeno) e 79 (dibutilftalato) (Figura 6) apresentaram maior potencial toxicológico com pEC₅₀ 5.56, 5.12 e 5.76, respectivamente. Os compostos mencionados apresentam valores elevados de D1, D2 e HSA, caracterizando moléculas hidrofóbicas. Em contraste, os compostos 12, 82 e 86 foram menos propensos a desenvolver efeitos toxicológicos no modelo analisado (Figura 2 e Tabela 2). O composto 8 é um composto líquido utilizado como intermediário na produção de tensoativos e plastificantes que é muito tóxico em ambientes aquáticos, tendo a capacidade de se acumular em organismos. O composto 35 mostra uma similaridade estrutural que difere apenas na posição de ligação do átomo de cloro no grupo aromático, sendo considerados isômeros de triclorobenzeno. Este é usado como intermediário em produtos químicos, particularmente em herbicidas e utilizado como solvente para dissolver materiais como óleos, ceras, resinas e graxas. Os clorobenzenos apresentam efeitos toxicológicos e grande capacidade de bioacumulação, são listados pela US. EPA como poluente prioritário (CHENG *et al.*, 2016; KIM and RYU, 2007). Em outro estudo, os animais submetidos à inalação deste composto apresentaram depressão da frequência respiratória (KORSAK *et al.*, 1995). O composto 79 tem sido amplamente utilizado como plastificantes em inúmeros produtos,

dentre estes, embalagens de alimentos, pavimentação, mobiliários, bem como presente em produtos de cuidado pessoal. É solúvel em vários solventes orgânicos, como álcool e benzeno. Sua exposição pode causar danos ao sistema endócrino causando tumores cancerosos, defeitos congênitos e outros distúrbios do desenvolvimento (BELTIFA, *et al.*, 2017; WILLIAMS *et al.*, 2016).

Os compostos 12 (terc-butil metil éter), 82 (nitrobenzeno) e 86 (4-nitrotolueno) apresentaram índices de baixa toxicidade (Figura 6), com pEC_{50} 2,19, 3,02 e 3,54, respectivamente. Ao contrário do composto 12, os compostos 82 e 86 mostram uma similaridade estrutural com os grupos nitroaromáticos. De acordo com Alves *et al.* (2016), os grupos nitro adicionados aos anéis aromáticos aumentam a toxicidade dos compostos; o que indica que a toxicidade variará de acordo com os substituintes adicionados.

Esses dados corroboram com os dados obtidos através da Figura 6, onde os MIFs podem ser visualizados. Os compostos mais tóxicos (8, 35 e 79) apresentam uma região hidrofóbica maior (zona verde), proporcionando-lhes uma característica mais lipofílica. Embora não haja regiões completamente cobertas pelo efeito hidrofóbico, o composto 79 é relativamente grande e simétrico apresentando características lipofílicas. Os compostos com baixa toxicidade (12, 82, 86) têm regiões hidrofóbicas menores, apresentando características mais polares, facilitando o potencial de ligação de hidrogênio, tornando as moléculas hidrófilas.

Figura 6 - MIFs: as regiões verdes representam uma sonda hidrofóbica em um nível de energia de $-0,2$ kcal/mol para compostos de maior toxicidade (8, 39, 79) e de menor toxicidade (12, 82, 86).



Não só os compostos mais hidrofóbicos caracterizados pelos descritores D1, D2 e ASH são mais tóxicos, mas o equilíbrio entre as regiões hidrofóbicas ao redor dos compostos é determinante para o MoA 1. É possível visualizar os compostos mais tóxicos, a região hidrofóbica (cor verde) cobre a maior parte da molécula (compostos 8 e 79, Figura 6) ou são distribuídos simetricamente em torno deles (composto 79). Por outro lado, os compostos menos tóxicos apresentam superfícies hidrofóbicas assimetricamente distribuídas em torno das moléculas (12, 82, 86).

4 CONCLUSÃO

Neste estudo, propusemos um modelo usando descritores do Volsurf para prever a toxicidade aguda dos peixes nos compostos MoA 1. Dos descritores selecionados, não apenas a hidrofobicidade está relacionada à toxicidade, mas outras propriedades físico-químicas que

combinadas contribuem para a atividade destes compostos, como o modelo PLS selecionado nos ajuda a encontrar. A distribuição simétrica das porções hidrofóbicas na estrutura dos compostos, bem como a forma, como cadeias ramificadas, são características importantes que estão relacionadas à toxicidade destes compostos. Esses modelos são novos e valioso para a redução de testes de toxicidade em peixes e avaliação risco de poluentes orgânicos no meio ambiente.

REFERÊNCIAS

- AFANTITIS, A. et al. Ligand - Based virtual screening procedure for the prediction and the identification of novel β -amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. **European Journal of Medicinal Chemistry**, v. 46, n. 2, p. 497–508, 2011.
- AHLERS, J.; NENDZA, M.; SCHWARTZ, D. Environmental hazard and risk assessment of thiochemicals. Application of integrated testing and intelligent assessment strategies (ITS) to fulfil the REACH requirements for aquatic toxicity. **Chemosphere**, v. 214, p. 480–490, 2019.
- ALEXANDER, D. L. J.; TROPSHA, A.; WINKLER, D. A. Beware of R^2 : Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. **Journal of Chemical Information and Modeling**, v. 55, n. 7, p. 1316–1322, 2015.
- ALTENBURGER, R. et al. Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. **Environmental Sciences Europe**, v. 31, n. 1, p. 1–17, 2019.
- ALVES, V. M. et al. Alarms about structural alerts. **Green Chemistry**, v. 18, n. 16, p. 4348–4360, 2016a.
- ALVES, V. M. et al. QSAR models of human data can enrich or replace LLNA testing for human skin sensitization. **Green Chemistry**, v. 18, n. 24, p. 6501–6515, 2016b.
- BARONI, M. et al. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. **Quantitative Structure-Activity Relationships**, v. 12, n. 1, p. 9–20, 1993.
- BELTIFA, A. et al. Plasticizers and bisphenol A, in packaged foods sold in the Tunisian markets: study of their acute in vivo toxicity and their environmental fate. **Environmental Science and Pollution Research**, v. 24, n. 28, p. 22382–22392, 2017.
- BOURGIN, M. et al. Differential chemical profiling to identify ozonation by-products of estrone-sulfate and first characterization of estrogenicity in generated drinking water. **Water Research**, v. 47, p. 3791–3802, 2013.
- BURDEN, N. et al. The utility of QSARs in predicting acute fish toxicity of pesticide metabolites: A retrospective validation approach. **Regulatory Toxicology and Pharmacology**, v. 80, p. 241–246, 2016.

CASSANI, S. et al. Daphnia and fish toxicity of (benzo)triazoles: Validated QSAR models, and interspecies quantitative activity-activity modelling. **Journal of Hazardous Materials**, v. 258–259, p. 50–60, 2013.

CASSOTTI, M. et al. A similarity-based QSAR model for predicting acute toxicity towards the fathead minnow (*Pimephales promelas*). **SAR and QSAR in Environmental Research**, v. 26, n. 3, p. 217–243, 2015.

CHEN, B. et al. Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: A review of methods and resources. **Journal of Hazardous Materials**, v. 299, p. 260–279, 2015.

CHENG, Z. et al. Synthesis, characterization, and photocatalytic activity of porous La–N–co-doped TiO₂ nanotubes for gaseous chlorobenzene oxidation. **Journal of Environmental Sciences (China)**, v. 46, p. 203–213, 2016.

CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 3, p. 682–692, 2002.

CRUCIANI, G. et al. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. **Journal of Molecular Structure: THEOCHEM**, v. 503, n. 1–2, p. 17–30, 2000.

DE OLIVEIRA, L. L. D. et al. Acute and chronic ecotoxicological effects of four pharmaceuticals drugs on cladoceran *Daphnia magna*. **Drug and Chemical Toxicology**, v. 39, n. 1, p. 13–21, 2016.

DELGADO, L. F. et al. QSAR-like models: A potential tool for the selection of PhACs and EDCs for monitoring purposes in drinking water treatment systems - A review. **Water Research**, v. 46, n. 19, p. 6196–6209, 2012.

ECHA. **The Use of Alternatives to Testing on Animals for the REACH Regulation**. [s.l.: s.n.]. v. 117

ERMONDI, G. et al. An application of two MIFs-based tools (Volsurf+ and Pentacle) to binary QSAR: The case of a palinurin-related data set of non-ATP competitive Glycogen Synthase Kinase 3 β (GSK-3 β) inhibitors. **European Journal of Medicinal Chemistry**, v. 46, n. 3, p. 860–869, 2011.

ESTEBAN, S. et al. Analysis and occurrence of endocrine-disrupting compounds and estrogenic activity in the surface waters of Central Spain. **Science of the Total Environment**, v. 466–467, p. 939–951, 2014.

GADALETA, D. et al. Applicability Domain for QSAR Models. **International Journal of Quantitative Structure-Property Relationships**, v. 1, n. 1, p. 45–63, 2016.

GASTEIGER, J. Chemoinformatics: Achievements and challenges, a personal view. **Molecules**, v. 21, n. 2, 2016.

GRAMATICA, P.; CASSANI, S.; SANGION, A. Aquatic ecotoxicity of personal care products: QSAR models and ranking for prioritization and safer alternatives' design. **Green Chemistry**, v. 18, n. 16, p. 4393–4406, 2016.

GROS, M. et al. Screening and prioritization of micropollutants in wastewaters from on-site sewage treatment facilities. **Journal of Hazardous Materials**, v. 328, p. 37–45, 2017.

GROS, M.; RODRÍGUEZ-MOZAZ, S.; BARCELÓ, D. Fast and comprehensive multi-residue analysis of a broad range of human and veterinary pharmaceuticals and some of their metabolites in surface and treated waters by ultra-high-performance liquid chromatography coupled to quadrupole-linear ion trap tandem. **Journal of Chromatography A**, v. 1248, p. 104–121, 2012.

HANSCH, C.; VERMA, R. P. Larvicidal activities of some organotin compounds on mosquito larvae: A QSAR study. **European Journal of Medicinal Chemistry**, v. 44, n. 1, p. 260–273, 2009.

HICKS, K. A. et al. Reduction of Intersex in a Wild Fish Population in Response to Major Municipal Wastewater Treatment Plant Upgrades. **Environmental Science and Technology**, v. 51, n. 3, p. 1811–1819, 2017.

ISIDORI, M. et al. Chemical and toxicological characterisation of anticancer drugs in hospital and municipal wastewaters from Slovenia and Spain. **Environmental Pollution**, v. 219, p. 275–287, 2016.

JURADO, A. et al. Emerging organic contaminants in groundwater in Spain: A review of sources, recent occurrence and fate in a European context. **Science of the Total Environment**, v. 440, p. 82–94, 2012.

KLÜVER, N. et al. Development of a general baseline toxicity QSAR model for the fish embryo acute toxicity test. **Chemosphere**, v. 164, p. 164–173, 2016.

KONEMANN, H. (QSARs). v. 19, 1981.

LEVET, A. et al. Quantitative structure-activity relationship to predict acute fish toxicity of organic solvents. **Chemosphere**, v. 93, n. 6, p. 1094–1103, 2013.

LUTTERBECK, C. A. et al. Evaluation of the toxic effects of four anti-cancer drugs in plant bioassays and its potency for screening in the context of waste water reuse for irrigation. **Chemosphere**, v. 135, p. 403–410, 2015.

MELAGRAKI, G. et al. In silico exploration for identifying structure-activity relationship of MEK inhibition and oral bioavailability for isothiazole derivatives. **Chemical Biology and Drug Design**, v. 76, n. 5, p. 397–406, 2010.

MIRANDA, A. C. et al. Surface water disinfection by chlorination and advanced oxidation processes: Inactivation of an antibiotic resistant E. coli strain and cytotoxicity evaluation. **Science of the Total Environment**, v. 554–555, p. 1–6, 2016.

MOUSEL, D.; PALMOWSKI, L.; PINNEKAMP, J. Energy demand for elimination of organic micropollutants in municipal wastewater treatment plants. **Science of the Total Environment**, v. 575, p. 1139–1149, 2017.

NENDZA, M.; MÜLLER, M.; WENZEL, A. Discriminating toxicant classes by mode of action: 4. Baseline and excess toxicity. **SAR and QSAR in Environmental Research**, v. 25, n. 5, p. 393–405, 2014.

NENDZA, M.; MÜLLER, M.; WENZEL, A. Classification of baseline toxicants for QSAR predictions to replace fish acute toxicity studies. **Environmental science. Processes & impacts**, v. 19, n. 3, p. 429–437, 2017.

OECD. Disponível em: <<https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm>>.

RIBEIRO, A. R. et al. An overview on the advanced oxidation processes applied for the treatment of water pollutants defined in the recently launched Directive 2013/39/EU. **Environment International**, v. 75, p. 33–51, 2015.

RICHTER, E. et al. Phytotoxicity of wastewater-born micropollutants - Characterisation of three antimycotics and a cationic surfactant. **Environmental Pollution**, v. 208, p. 512–522, 2016.

ROY, K.; AMBURE, P. The “double cross-validation” software tool for MLR QSAR model development. **Chemometrics and Intelligent Laboratory Systems**, v. 159, p. 108–126, dez. 2016.

ROY, K.; KAR, S.; AMBURE, P. On a simple approach for determining applicability domain of QSAR models. **Chemometrics and Intelligent Laboratory Systems**, v. 145, p. 22–29, 2015.

SCHLOSSER, P. M.; BOGDANFFY, M. S. Determining modes of action for biologically based risk assessments. **Regulatory Toxicology and Pharmacology**, v. 30, n. 1, p. 75–79, 1999.

SCHÜÜRMAN, G.; EBERT, R. U.; KÜHNE, R. Quantitative read-across for predicting the acute fish toxicity of organic compounds. **Environmental Science and Technology**, v. 45, n. 10, p. 4616–4622, 2011.

SERRA-ROIG, M. P. et al. Occurrence, fate and risk assessment of personal care products in river–groundwater interface. **Science of the Total Environment**, v. 568, p. 829–837, 2016.

SINGH, K. P. et al. Multispecies QSAR modeling for predicting the aquatic toxicity of diverse organic chemicals for regulatory toxicology. **Chemical Research in Toxicology**, v. 27, n. 5, p. 741–753, 2014.

TAMURA, I. et al. Contribution of pharmaceuticals and personal care products (PPCPs) to whole toxicity of water samples collected in effluent-dominated urban streams. **Ecotoxicology and Environmental Safety**, v. 144, n. July 2016, p. 338–350, 2017.

THOMAS, P. et al. Application of the Activity Framework for Assessing Aquatic Ecotoxicology Data for Organic Chemicals. **Environmental Science and Technology**, v. 49, n. 20, p. 12289–12296, 2015.

TODESCHINI, R. **Molecular Descriptors for Chemoinformatics, Volumes I & II**. [s.l.: s.n.].

TOROPOVA, A. P. et al. CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna*. **Chemometrics and Intelligent Laboratory Systems**, v. 110, n. 1, p. 177–181, 2012.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, 12 jul. 2010.

TUGCU, G.; ERTÜRK, M. D.; SAÇAN, M. T. On the aquatic toxicity of substituted phenols to *Chlorella vulgaris*: QSTR with an extended novel data set and interspecies models. **Journal of Hazardous Materials**, v. 339, p. 122–130, 2017.

VERHAAR, H. J. M. et al. Classifying environmental pollutants: Part 3. External validation of the classification system. **Chemosphere**, v. 40, n. 8, p. 875–883, 2000.

WILLIAMS, M. J. et al. Dibutyl Phthalate Exposure Disrupts Evolutionarily Conserved Insulin and Glucagon-Like Signaling in *Drosophila* Males. **Endocrinology**, v. 157, n. 6, p. 2309–2321, 1 jun. 2016.

WOLD, S. et al. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. **SIAM Journal on Scientific and Statistical Computing**, v. 5, n. 3, p. 735–743, set. 1984.

WOLD, S. et al. The PLS model space revisited. **Journal of Chemometrics**, v. 23, n. 2, p. 67–68, 2009.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

XUE, X. et al. Trophic magnification of parabens and their metabolites in a subtropical marine food web. **Environmental Science and Technology**, v. 51, n. 2, p. 780–789, 2017.

CAPÍTULO III

CAPÍTULO III - FERRAMENTAS COMPUTACIONAIS PREDITIVAS PARA AVALIAÇÃO DE ECOTOXICIDADE DE MICROPOLUENTES ORGÂNICOS ENCONTRADOS EM FONTES AQUÁTICAS DO BRASIL ²

RESUMO

O aumento de poluentes químicos detectados em diferentes ambientes aquáticos nos últimos anos tem sido foco de vários estudos relacionados à sua ocorrência, transporte, destino e riscos à saúde humana e ambiental. No Brasil, estudos recentes têm sido conduzidos sobre a ocorrência de uma série de micropoluentes orgânicos (MPO) em ambientes aquáticos. No entanto, a informação toxicológica e o comportamento ambiental para a maioria desses poluentes ainda são difíceis de avaliar. Obtivemos modelos de consenso utilizando uma combinação dos descritores VolSurf e Dragon e a abordagem Random Forest para prever a toxicidade de compostos em *Daphnia magna* e *Pimephales promelas* para compostos identificados em diferentes fontes de água no Brasil. Os modelos mostraram desempenho satisfatório quando comparados com outras abordagens, como o algoritmo T.E.S.T (do inglês, *Toxicity Estimation Software Tool*), etc. Ambos os modelos podem ser usados como ferramentas complementares para auxiliar na previsão de compostos orgânicos contra ambos os organismos.

Palavras-chave: Micropoluentes orgânicos, Ecotoxicidade, *Daphnia magna*, *Pimephales promelas*, modelos QSAR.

² Capítulo publicado na revista *Molecular Informatics*.

ABSTRACT

The increase of chemical pollutants detected in different aquatic environments over the past few years has been in the focus of several studies related to their occurrence, transport, fate, and hazards, or risks to human and environmental health. In Brazil, recent studies have been conducted on the occurrence of a series of organic micropollutants in aquatic environments. Nevertheless, the toxicological information and environmental behavior for most of these pollutants are still difficult to evaluate. We obtained consensus models using a combination of VolSurf and Dragon descriptors and Random Forest approach to predict the toxicity against *Daphnia magna* and *Pimephales promelas* for compounds identified in different sources of water in Brazil. The models showed satisfactory performance when compared with other approaches, such as Toxicity Estimation Software Tool (T.E.S.T.), etc. Both models can be used as complementary tools to aid in the prediction of organic compounds against both organisms.

Keywords: Organic micropollutants, *Ecotoxicity*, *Daphnia magna*, *Pimephales promelas*, *QSAR models*.

1 INTRODUÇÃO

O aumento de poluentes químicos detectados em diferentes ambientes aquáticos nos últimos anos foi foco de diversos estudos relacionados à ocorrência, transporte, destino e riscos à saúde humana e ambiental (GAVRILESCU et al., 2015; HAMID; LI; GRACE, 2018; PETRIE; BARDEN; KASPRZYK-HORDERN, 2015). Micropoluentes orgânicos (MPOs), também chamados de contaminantes emergentes, incluem uma ampla gama de substâncias químicas extensivamente utilizadas em residências e aplicações industriais, tais como produtos farmacêuticos e de cuidados pessoais (BIELEN et al., 2017; EBELE; ABOU-ELWafa ABDALLAH; HARRAD, 2017; MANDARIC et al., 2017), desreguladores endócrinos (MELO; BRITO, 2014), pesticidas (MUNZ et al., 2017), tensoativos (FEDEILA et al., 2018), nanomateriais (SANTHOSH et al., 2016), metabólitos e seus produtos de transformação (DEEB et al., 2017).

A principal fonte de MPO para o meio ambiente tem sido fortemente atribuída ao lançamento de efluente resultante das estações de tratamento convencionais, uma vez que as técnicas mais frequentemente utilizadas para o tratamento das águas residuais não são especificamente projetadas para remover poluentes que ocorrem tipicamente em baixas concentrações na ordem de nanogramas por litro (ng.L^{-1}) a microgramas por litro ($\mu\text{g.L}^{-1}$) (BUI et al., 2016; GROS et al., 2017; LUO et al., 2014). Outras fontes de micro poluentes também podem resultar do escoamento agrícola e pecuária (SUN et al., 2018), aquicultura (TAVAKOL et al., 2017), bem como efluentes hospitalares (WANG; WANG; YANG, 2018). Consequentemente, isso leva à detecção de muitos micropoluentes orgânicos em diferentes fontes aquáticas, como águas superficiais (NEALE et al., 2017), águas subterrâneas (AN et al., 2018), águas oceânicas (NÖDLER; VOUTSA; LICHA, 2014) e até mesmo água potável (YANG et al., 2017).

Embora esses MPO ocorram em baixas concentrações, esses são continuamente liberados no meio aquático com propriedades físico-químicas particulares, muitas vezes reconhecidas por propriedades indesejáveis nas quais podem exercer efeitos nocivos sobre o ecossistema e a saúde humana (ARNOT *et al.*, 2008; EBELE *et al.*, 2017). Alguns produtos são caracterizados como sendo compostos farmacêuticos quimicamente ativos, resistentes à degradação e altamente persistentes em meios aquosos que apresentam um risco potencial de

toxicidade quando no meio ambiente (DESBIOLLES et al., 2018; RIVERA-UTRILLA et al., 2013). Os produtos químicos estrogênicos podem interferir significativamente na função do sistema endócrino na vida selvagem e nos seres humanos (BARREIROS et al., 2016), associado às funções de desenvolvimento e reprodução, particularmente em organismos aquáticos (KOBAYASHI et al., 2017). Além disso, a ocorrência e a diversidade de antibióticos foram relatadas, facilitando o desenvolvimento e disseminação da resistência a antibióticos em bactérias (JIA et al., 2018). Muitos antibióticos e seus produtos de transformação podem direta ou indiretamente apresentar genotoxicidade para várias linhagens celulares (TOOLARAM et al., 2016). Além disso, os poluentes orgânicos hidrofóbicos são compostos com tendência natural para evitar a fase aquosa quando introduzidos em ambientes aquáticos, com elevada resistência à biotransformação e biodegradação, e conhecidos pelo seu potencial de bioacumulação em organismos através da transferência de produtos químicos ao longo da teia alimentar.

Recentemente, LIU et al. relataram um acúmulo de poluentes orgânicos persistentes em diversos animais selvagens aquáticos e terrestres, indicando potencial de bioacumulação dependente do habitat entre diferentes espécies e alta correlação entre o potencial de transferência materna de poluentes orgânicos halogenados (LIU et al., 2018). Além disso, os pesticidas também podem ter efeitos adversos nos sistemas aquáticos, estes abrangem uma vasta gama de compostos e produtos químicos e refere-se a uma das categorias mais comuns de poluentes que podem ser persistentes em água ou bioacumuláveis na biota, dependendo da sua solubilidade e do coeficiente de partição octanol-água (CCANCCAPA et al., 2016; HEYS et al., 2016). Um dos principais problemas relacionados ao uso destes compostos é que eles não são totalmente seletivos e, portanto, podem causar efeitos tóxicos em organismos não-alvo, representando uma ameaça ao funcionamento normal dos ecossistemas aquáticos e terrestres (MANSANO et al., 2016). Para proteger a saúde humana e ambiental da poluição química no ambiente aquático é necessário ter conhecimento preciso sobre o poluente e entender o mecanismo pelo qual o químico pode causar riscos e impactos na saúde de um ecossistema e na qualidade da água (YOU et al., 2015). Embora amplamente detectada no ambiente, a maioria dos micropoluentes é mal inventariada e regulada, sem ou com insuficiente informação relacionada ao seu impacto na qualidade da água (superficial, subterrânea ou de consumo). Algumas ações inovadoras nessa direção têm sido

desenvolvidas por agências reguladoras como a Agência de Proteção Ambiental dos Estados Unidos (do inglês, *United States Environmental Protection Agency* – U.S. EPA) e Agência Europeia das Substâncias Químicas a partir do regulamento de Registro, Avaliação, Autorização e Restrição de Produtos Químicos (do inglês, *Registration, Evaluation, Authorisation and Restriction of Chemicals* - REACH) (BARBOSA et al., 2016).

Para estabelecer limites regulatórios para MPOs, pesquisas adicionais sobre respostas biológicas desses compostos são de particular importância. Assim, para um estudo de avaliação de risco uma melhor compreensão da sua toxicidade e ecotoxicidade é essencial e um pré-requisito para uma proteção abrangente do meio ambiente (MONTAGNER, *et al.*, 2017; OECD, 2014). Métodos computacionais baseados em estudo matemático de relações quantitativas entre estrutura química e a atividade/propriedade de um composto [do inglês (*Quantitative*) *Structure-Activity Relationships* - (Q)SARs] desempenham um papel importante no campo da química medicinal (CAPUZZI et al., 2018), design de drogas (ZAKHAROV et al., 2016) e toxicologia (GRAMATICA; CASSANI; SANGION, 2016), tornando-se cada vez mais aplicada nas ciências ambientais como ferramentas de triagem para classificar e priorizar produtos químicos que podem apresentar impacto adverso na saúde humana e ambiental (GRAMATICA; PAPA; SANGION, 2018). O QSAR é derivado da aplicação de ferramentas matemáticas e estatísticas, resultando em uma correlação entre propriedades físico-químicas e atividades biológicas, numericamente codificadas por um ou mais descritores moleculares que são usados para prever as propriedades de interesse (CONSONNI; TODESCHINI; PAVAN, 2002; DEARDEN; ROTUREAU; FAYET, 2013; GASTEIGER, 2016; TODESCHINI, 2009).

A aplicação de tais métodos é uma alternativa útil para lidar com grandes quantidades de dados, avaliar as propriedades das substâncias, seus processos determinantes de destino e prever seus efeitos, reduzindo experimentos caros e demorados e mostrando um forte potencial para reduzir o uso de animais utilizados em ensaios experimentais (DE MORAIS E SILVA et al., 2018; NICOLAS et al., 2018). Em um contexto de tomada de decisão, a proteção abrangente da saúde pública e do meio ambiente com o uso de métodos preditivos computacionais apoia os esforços contínuos para desenvolver informações a fim de preencher as lacunas nos dados de eco(toxicidade), necessário para melhorar a compreensão da avaliação dos perigos dos produtos químicos conforme exigido por várias agências

governamentais em todo o mundo, incluindo a U.S. EPA e regulamentos adotados pela UE através do REACH (EBELE; ABOU-ELWAFAB ABDALLAH; HARRAD, 2017; TODESCHINI, 2009).

No Brasil, estudos recentes têm sido realizados a fim de investigar a ocorrência de uma série de MPO em ambientes aquáticos. No entanto, as informações toxicológicas para a maioria destes poluentes ainda são difíceis de avaliar. Portanto, no presente estudo, examinamos os dados coletados da literatura disponível on-line (Web of Science) que reportam a presença de MPO, de diferentes classes químicas, nas matrizes de águas residuais, superficiais, subterrâneas e de água potável no Brasil (DOS PASSOS MENEZES et al., 2016; LOCATELLI; SODRÉ; JARDIM, 2011; MACHADO et al., 2016; SODRÉ et al., 2010). O principal objetivo foi desenvolver um modelo preditivo de ecotoxicidade para MPO identificados em diferentes fontes de água no Brasil para crustáceo de água doce (*Daphnia magna*) e peixe (*Pimephales promelas*), ambos comumente usados em estudos de toxicidade aguda. Além disso, os modelos gerados também foram comparados em relação ao desempenho preditivo para verificar a contribuição útil para a saúde humana e o meio ambiente da redução dos riscos impostos pelos produtos químicos.

2 MATERIAL E MÉTODOS

2.1 Conjuntos de dados

Os dados experimentais foram obtidos da plataforma OCHEM (do inglês, *Online CHEmical Modeling environment*). O conjunto de dados incluiu uma ampla variedade de informações de toxicidade aquática de vários compostos orgânicos, os quais foram selecionados para os organismos específicos, tipo de atividade e tempo duração do teste. O conjunto de dados de *D. magna* incluiu 2759 registros químicos com estrutura química definida. Após a triagem de dados, que consiste na unificação dos dados químicos, incluindo remoção de inorgânicos, organometálicos, contraíons, sal e misturas, padronização de quimiotipos específicos e análise de duplicatas o conjunto final resultou em 993 produtos químicos orgânicos. Os dados de testes de imobilização aguda e os valores de ecotoxicidade experimentais foram expressos em CE₅₀ (concentração da amostra que causa efeito tóxico a

50% dos organismos) no tempo de exposição de 48h, seguindo a Organização para Cooperação Econômica e Desenvolvimento (OCDE) protocolos (OCDE, 202). A toxicidade aguda para peixes incluiu 1132 registros químicos com estrutura química definida. Após a triagem de dados, o conjunto de dados resultou em 1016 produtos químicos orgânicos com base em resultados de concentração letal de 50% (CL_{50} - concentração da amostra que causa mortalidade de 50% dos organismos) em 96 h para *Pimephales promelas* (teste OECD 203). Todos os valores de ecotoxicidade experimentais foram expressos em uma unidade molar (mol.L^{-1}) e, em seguida, os dados transformados logaritmicamente [pEC_{50} e pLC_{50}] foram usados como variáveis de resposta.

Todas as estruturas químicas do conjunto de dados original foram revisadas utilizando informações disponíveis no PubChem Compound (<http://www.ncbi.nlm.nih.gov/pccompound>), ChemSpider (<http://www.chemspider.com/>) ou no banco de dados CHRIP (https://www.nite.go.jp/en/chem/chrip/chrip_search/systemTOP) utilizando o número de registro CAS (do inglês, *Chemical Abstract Services*) e nome químico. Produtos químicos com estrutura química indefinida ou ausente não foram considerados para modelagem. Os conjuntos de dados foram cuidadosamente selecionados de acordo com o fluxo de trabalho desenvolvidos pelo nosso grupo (FOURCHES; MURATOV; TROPSHA, 2010, 2015, 2016). Resumidamente, a normalização estrutural de quimiotipos específicos, tais como grupos aromáticos e nitro, foi realizada usando o software ChemAxon Standardizer (JChem, 16.1.11.0, 2016, ChemAxon, Budapeste, Hungria, <http://www.chemaxon.com>). Sais inorgânicos, compostos organometálicos e misturas também foram removidos. Após a padronização estrutural, as duplicatas foram identificadas usando o HiT QSAR (MURATOV et al., 2010). JChem também foi usado para gerar estruturas tridimensionais (3D).

2.2 Desenvolvimento e validação de dados

A informação química das diferentes representações moleculares foi extraída através do cálculo de descritores moleculares que são variáveis numéricas que quantificam a informação estrutural de um produto químico (FOURCHES; MURATOV; TROPSHA, 2015). No presente trabalho, VolSurf + v. 1.0.7 (<http://www.moldiscovery.com>) e Dragon

v.7.04 foram usados para o cálculo do descritor (CONSONNI; BALLABIO; TODESCHINI, 2009; CRUCIANI; PASTOR; GUBA, 2000). Os modelos QSAR foram desenvolvidos e rigorosamente validados de acordo com as melhores práticas de modelagem QSAR (CHERKASOV et al., 2014; FOURCHES; MURATOV; TROPSHA, 2010). Antes da modelagem, estimamos o índice de modelabilidade (GOLBRAIKH et al., 2014) para cada conjunto de dados. Os conjuntos de dados foram divididos aleatoriamente em um conjunto de treinamento e teste, abrangendo 80% e 20% dos compostos, respectivamente.

Os modelos foram gerados usando o conjunto de treinamento e nós de aprendizados supervisionados Random Forest (RF) para os descritores Dragon e VolSurf, totalizando dois modelos de RF para cada ecotoxicidade (BREIMAN, 2001). Um modelo de consenso (MC) foi obtido a partir do cálculo dos valores médios dos dois modelos gerados pela combinação de Random Forest com os descritores Dragon e VolSurf, para cada atividade, afim de maximizar a precisão da predição dos modelos. A predição da toxicidade foi estimada adotando uma média das toxicidades previstas dos métodos QSAR (desde que as previsões estejam dentro dos respectivos domínios de aplicabilidade – DA). Esse método geralmente fornece melhor precisão no modelo de predição à maioria dos modelos individuais, pois as previsões errôneas são atenuadas pelas previsões dos outros métodos.

O DA baseado nas distâncias euclidianas, empregando o nó Enalos no software KINME 3.14.1 foi usado para identificar compostos no conjunto de testes para os quais as previsões podem não ser confiáveis se os valores forem maiores que $DA = d + Z\sigma$, onde d e σ são média da distância euclidiana e o desvio padrão do conjunto de amostras do conjunto de treinamento que possuem valores médios de distância Euclidiana mais baixos de todas as amostras no conjunto de treinamento. O parâmetro Z é um valor de corte empírico, 0,5 foi usado como padrão (TROPSHA; GOLBRAIKH, 2007).

No final, após obter e validar os modelos, estes foram comparados com modelos de QSAR disponíveis na plataforma VEGA (www.vegahub.eu/), CAESAR (www.caesar-project.eu/) e T.E.S.T. (www.epa.gov/chemical-research/toxicity-estimation-software-tool-test). Este último inclui as seguintes abordagens: DEMETRA, EPA, SARpy e toxRead.

3 RESULTADOS E DISCUSSÕES

3.1 Desenvolvimento e Validação dos Modelos

O modelo preditivo dividido em treino e teste, para *D. magna* e *P. promelas*, é apresentado na Tabela 1. Aproximadamente 80% dos compostos participaram do conjunto de treinamento e o restante do conjunto original, 20%, foram separados para o conjunto de teste.

Tabela 1 - Conjunto de dados de *D. magna* e *P. promelas*

	Conjunto treinamento	%Total	Conjunto teste	%Total
<i>D. magna</i>	794	79.96	199	20.04
<i>P. promelas</i>	803	79.04	213	20.96

Os modelos de predição de ecotoxicidade para *P. promelas* (Tabela 2) demonstraram desempenho significativo na validação cruzada ($R^2 = 0,707$ para os descritores VolSurf (VS) e $R^2 = 0,731$ para os descritores Dragon (DG)). No conjunto de testes, os descritores de DG mostraram uma resposta de predição semelhante ($R^2 = 0,723$) aos descritores de VS ($R^2 = 0,713$). A abordagem consensual mostrou melhor desempenho de predição em comparação com o modelo individual com $R^2 = 0,740$ para validação cruzada e $R^2 = 0,743$ para o conjunto de teste.

A qualidade dos modelos para *D. magna* (Tabela 2) foi significativamente menor. Os descritores Dragon e VolSurf não alcançaram o valor de R^2 de 0,6, embora o desempenho do modelo de consenso no conjunto de testes tenha sido muito próximo ($R_{ext}^2 = 0,592$). Quadro semelhante foi observado para a métrica, coeficiente de correlação de validação cruzada externa (Q^2) (ver Tabela 3).

Tabela 2 - Coeficiente de determinação de validação cruzada R^2 de 10 vezes e conjuntos de testes (R_{ext}^2) para os modelos Dragon, VolSurf e Consensus Random Forest

Descritores	<i>Daphnia magna</i>		<i>Pimephales promelas</i>	
	R^2	R_{ext}^2	R^2	R_{ext}^2
Dragon	0.534	0.583	0.731	0.723

Volsurf	0.514	0.521	0.707	0.713
Consenso	0.563	0.592	0.740	0.743

Tabela 3 - O coeficiente de correlação por validação cruzada (Q_{cv}^2) e a variância explicada na predição (Q_{ext}^2) para os modelos Dragon, VolSurf e Consensus Random Forest

Descritores	<i>Daphnia magna</i>		<i>Pimephales promelas</i>	
	Q_{cv}^2	Q_{ext}^2	Q_{cv}^2	Q_{ext}^2
Dragon	0.524	0.500	0.728	0.788
Volsurf	0.513	0.451	0.705	0.783
Consenso	0.547	0.507	0.737	0.804

3.2 Desempenho de diferentes abordagens

Os CMs desenvolvidos neste estudo foram usados para prever um conjunto de 64 micropoluentes orgânicos de diferentes classes encontrados nas fontes de águas residuais, superficiais, subterrâneas e de água potável no Brasil. Além disso, o desempenho preditivo foi comparado com três algoritmos para cada organismo. Para *D. magna*, foram utilizados os seguintes algoritmos: O EPA versão 1.0.7, um modelo (Q) SAR para *D. magna* LC₅₀ (48h), baseado em regressão linear múltipla e DEMETRA, versão 1.0.4: toxicidade aguda para *D. magna* com pesticidas LC₅₀ (48h) construída como modelo híbrido, utilizando duas Redes Neurais Artificiais (RNA) e um Quadrado Mínimo Parcial (PLS).

Para o *P. promelas*, os seguintes algoritmos foram utilizados para prever a toxicidade de 64 MPO: ToxRead versão 1.0.0 que realizou uma avaliação para toxicidade aguda integrando a um conjunto de dados de 972 produtos químicos. O modelo read-across foi construído com o aplicativo istKNN (desenvolvido pelo Kode) e é baseado no índice de similaridade (MANGANARO *et al.*, 2016). O SARpy (SAR em python) é baseado em um modelo SAR (relação estrutura-atividade). A predição do software T.E.S.T., predisse os valores de LC₅₀ (48h) para *D. magna* e LC₅₀ (96h) para *P. promelas*.

3.2.1 Modelagem de MPO para *Daphnia magna*

Inicialmente, a concordância entre os modelos foi analisada usando as matrizes de correlação da predição. Para *D. magna*, todos os métodos geraram valores de predição para todos os compostos, exceto T.E.S.T. que gerou valores de LC₅₀ para 54 compostos. Em geral, obtém-se uma melhor performance de predição utilizando o método de consenso uma vez que este é obtido a partir da média dos dados de todos os modelos de predição (Tabela 4). Para o conjunto de teste, usamos critérios menos restritivos para CM, porque, se uma amostra é confiável de acordo com a aplicabilidade do domínio utilizada, o valor é gerado. O domínio de aplicabilidade do modelo VS RF e dos modelos DG RF foi confiável, exceto para um composto e dois compostos, respectivamente, pois estavam fora da faixa de limite do D.A.. Portanto, os valores de LC₅₀ foram gerados para todos os 64 compostos. As abordagens CM e o T.E.S.T. mostraram uma alta correlação ($r = 0,887$) para os 54 compostos previstos (Tabela 5). O modelo CM mostrou uma correlação ainda melhor com os três métodos, EPA, DEMETRA e T.E.S.T.

Tabela 4 - Valores experimentais e previstos de pLC₅₀ de *D. magna* obtidos pelas quatro diferentes abordagens

ID	CM	DEMETRA	EPA	T.E.S.T.	Experimental
1	3.12	2.85	3.05	3.20	-
2	3.72	3.69	4.63	4.22	-
3	3.82	4.38	5.93	4.02	-
4	3.84	4.34	5.66	4.29	-
5	4.18	4.50	3.18	3.66	-
6	6.01	6.26	5.37	6.14	-
7	4.11	5.75	5.53	3.83	-
8	3.96	2.24	0.42	5.03	-
9	4.17	4.09	4.21	3.99	-
10	3.98	4.73	3.83	-	-
11	4.01	2.36	1.68	4.71	6.43
12	4.94	4.74	6.04	5.06	4.23

13	5.22	4.56	5.37	4.43	-
14	3.61	3.93	4.09	3.25	-
15	3.48	3.83	4.43	3.46	-
16	4.32	3.91	5.18	4.24	3.60
17	4.82	4.23	5.48	4.72	4.12
18	4.53	3.86	4.77	4.55	4.36
19	4.63	4.07	5.77	4.80	-
20	4.32	3.79	5.34	4.33	3.79
21	5.26	5.50	5.34	5.15	-
22	4.37	4.50	4.98	3.82	3.90
23	3.85	3.65	4.66	-	-
24	4c.91	6.10	2.03	-	-
25	5.20	5.17	5.40	4.33	-
26	4.64	5.37	3.82	-	-
27	5.12	4.56	6.13	4.96	4.82
28	5.29	4.39	5.84	4.61	4.35
29	3.34	3.71	5.34	-	-
30	4.92	4.09	6.18	-	-
31	5.94	4.30	6.78	-	-
32	3.68	3.55	5.24	4.04	-
33	6.27	4.68	7.94	-	-
34	3.55	3.31	4.90	3.72	-
35	5.07	3.93	4.22	4.93	4.96
36	5.85	6.56	6.65	5.89	-
37	5.69	3.66	4.20	5.10	5.37
38	4.92	3.73	2.96	4.70	-
39	4.82	3.91	5.56	5.05	-
40	5.12	5.18	5.93	-	-
41	5.58	5.39	6.04	-	-
42	5.37	5.17	6.32	4.44	-
43	5.96	6.73	6.93	5.99	-

44	4.90	4.67	5.66	4.59	5.17
45	4.67	5.38	6.24	4.29	-
46	6.65	6.21	6.57	5.55	-
47	4.90	4.15	5.05	4.48	-
48	4.11	5.18	3.73	4.27	-
49	6.76	6.95	6.40	6.90	-
50	5.56	5.62	4.64	4.22	5.30
51	7.15	8.12	6.34	6.38	6.45
52	6.78	6.77	6.48	6.40	6.65
53	5.85	6.82	6.03	5.36	6.51
54	5.59	3.79	4.66	5.01	-
55	8.27	8.98	7.01	8.20	7.99
56	4.04	4.21	5.18	4.10	-
57	4.88	5.90	5.46	4.57	-
58	6.28	6.05	5.63	4.82	-
59	4.96	5.96	4.48	4.53	4.71
60	3.65	4.03	3.65	3.94	-
61	3.59	4.00	3.75	3.61	-
62	4.98	6.60	4.24	5.09	4.58
63	4.74	4.62	3.65	4.82	-
64	5.10	5.06	3.80	4.38	5.22

Tabela 5 - Matriz de correlação (r) dos valores previstos de LC₅₀ de *D. magna* pelos quatro métodos utilizados neste estudo para 64 MPO identificados em diferentes matrizes de água no Brasil. a) Valores de LC₅₀ preditos apenas para 54 compostos

Modelos	CM	DEMETRA	EPA	T.E.S.T.^a
CM	1	0.764	0.556	0.887
DEMETRIA		1	0.503	0.707
EPA			1	0.458
T.E.S.T.				1

EPA, DEMETRA e T.E.S.T. forneceram alguns valores experimentais de ecotoxicidade para 20 compostos que foram usados para aplicar os critérios de validação propostos por Tropsha *et al.* (GOLBRAIKH; TROPSHA, 2002; TROPSHA; GRAMATICA; GOMBAR, 2003). Nesses critérios, os modelos foram considerados aceitáveis se satisfizessem todas as seguintes condições: $Q^2 > 0,5$, $R^2 > 0,6$, R_0^2 ou $R_0'^2$ próximo de R^2 , ou seja, e os correspondentes $0,85 \leq k \leq 1,15$ ou $0,85 \leq k' \leq 1,15$. Além disso, usamos o coeficiente modificado r_m^2 proposto por Roy^[79] para validação externa. Um $r_m^2 > 0.5$ indica uma boa previsibilidade externa. As abordagens CM e T.E.S.T geraram valores de R^2 superiores a 0,6 (Tabela 6) e baixo RMSE (do inglês, *Root Mean Square Error*) que reduziu o erro global do modelo. Além disso, de acordo com os critérios propostos por Tropsha e Roy (ROY; ROY, 2008; TROPSHA; GRAMATICA; GOMBAR, 2003) ambos os modelos mostraram previsões similares e satisfatórias, com o software T.E.S.T. gerando os melhores resultados (Tabela 7 e Figura 1).

Tabela 6 - Coeficiente de determinação e RMSE dos valores previstos de *D. magna* pLC₅₀ de 20 compostos de MPO identificados em diferentes matrizes de água no Brasil

Modelos	CM	DEMETRIA	EPA	T.E.S.T.
R_{ext}^2	0.626	0.388	0.0274	0.661
RMSE	0.720	1.255	1.481	0.665

Figura 1 - Gráfico de regressão dos valores de pLC₅₀ predito versus observados para *D. magna* (a e c) e valores observados versus preditos de pLC₅₀ (b e d) para consenso *Random Forest* (a e b) e software T.E.S.T. (c e d).

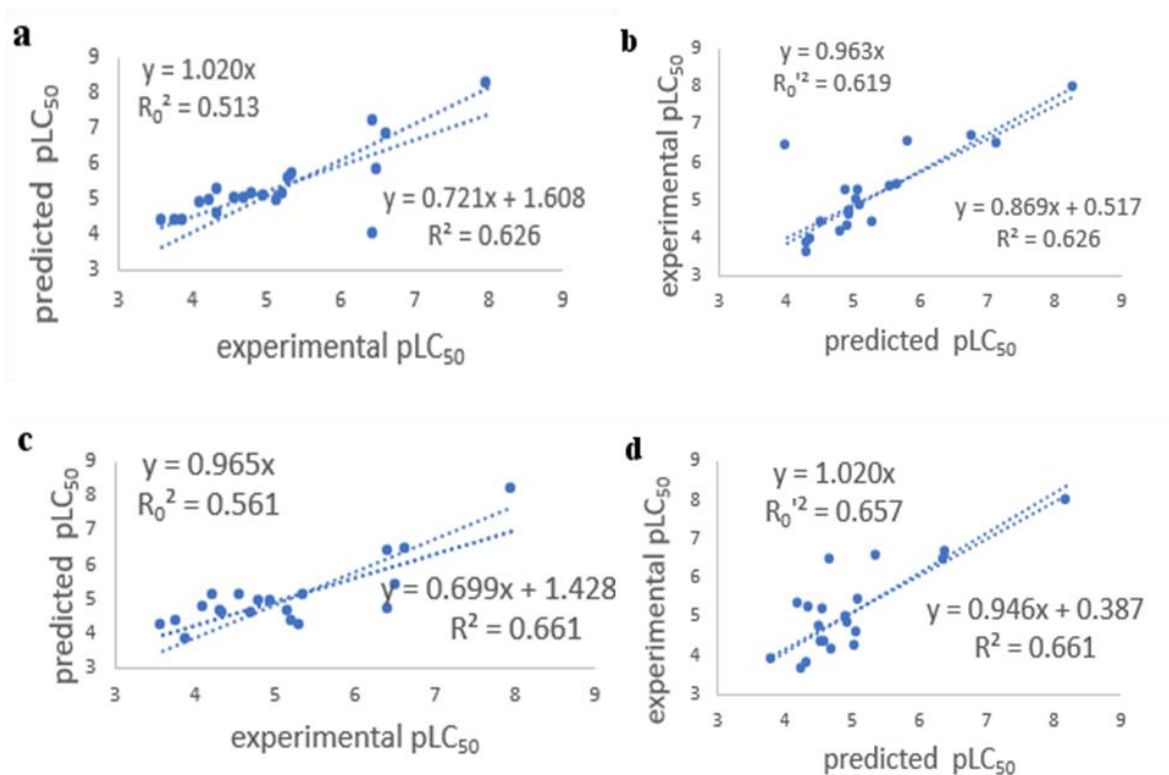


Tabela 7 - Os valores dos parâmetros (em negrito estão dentro da faixa permitida) utilizados pelo critério de validação do modelo, proposto por Tropsha e Roy, de CM e T.E.S.T. modelos para valores de *D. magna* pLC₅₀

Modelos	CM	T.E.S.T.
Rext^{2a}	0.626	0.661
Rext ₀ ^{2 b}	0.513	0.561
(Rext ² - Rext ₀ ²) / Rext ^{2 c}	0.181	0.151
Rext ₀ ^{'2 d}	0.619	0.657
(Rext ² - Rext ₀ ^{'2}) / Rext ^{2 e}	0.011	0.006
K^f	1.020	0.965
k'^g	0.963	1.020
r^{2m}h	0.416	0.452

$r^2_{m'}^i$	0.574	0.619
$\overline{r_m^2}^j$	0.495	0.535
$\overline{\Delta_m}^k$	0.158	0.167

^a valores do coeficiente de determinação (ajuste externo) entre os valores de pLC₅₀ observado e predito;

^b valores de coeficiente de determinação (ajuste externo) da linha cujo valor do intercepto y, entre os valores do pLC₅₀ preditos e observados, é igual a 0, obtidos dos gráficos a e c, Figura 1;

^c valor da diferença relativa entre os valores do coeficiente de determinação obtidos dos gráficos a e c, Figura 1;

^d valores do coeficiente de determinação (ajuste externo) da linha cujo valor do intercepto-y, entre os valores observados e preditos do pLC₅₀, é igual a 0, obtido dos gráficos b e d, Figura 1;

^e valor da diferença relativa entre os valores do coeficiente de determinação obtidos dos gráfico b e d, Figura 1;

^f declive da linha cujo intercepto y é igual a 0, mostrado nos gráficos a e c, Figura 1;

^g declive da linha cujo intercepto y é igual a 0, mostrado nos gráficos b e d, Figura 1;

^h métrica r^2_m , proposto por Roy (2008) para validação externa;

ⁱ métrica $r^2_{m'}$, proposto por Roy (2008) para validação externa utilizando R_{ext0} ';

^j average r^2_m , $\overline{r_m^2} = \frac{r_m^2 + r_{m'}^2}{2}$

^k $\overline{\Delta_m} = \frac{r_m^2 - r_{m'}^2}{2}$;

Entre os 64 MPO, os compostos 11, 49, 51, 52 e 55 foram preditos como mais tóxicos (pLC₅₀ > 6,34) para todos os modelos. Isso corroborou com valores experimentais de pLC₅₀ dos compostos 51, 52 e 55 (pLC₅₀ > 6,45). A toxicidade do composto 11 (Tabela 4 e Tabela 8) foi subestimada (pLC₅₀ < 4,71), apresentando valor muito inferior ao seu pLC₅₀ experimental = 6,43.

3.2.2 Modelagem de MPO para *Pimephales promelas*

A mesma análise utilizada para *D. magna* foi realizada para o modelo de *P. promelas*. O CM previu valores de pLC₅₀ para 46 compostos que estavam dentro do seu domínio de aplicabilidade dos modelos VS ou DG RF. O algoritmo Toxicity Read-Across versão 1.0.0 previu LC₅₀ para 45 compostos e T.E.S.T. para 54 compostos. Comparando com o SARpy que qualitativamente prediz a toxicidade em intervalos, classificando os compostos de 1 (LC₅₀ < 1 mg / L) até não-tóxico (LC₅₀ > 100 mg / L). As classes 2 e 3 correspondem aos intervalos entre 1 a 10 mg / L e 10 a 100 mg / L, respectivamente. Além disso, o modelo SARpy fornece a confiabilidade da previsão (Tabela 9).

Dos valores de atividades LC₅₀ previstos para 46 compostos pelo CM, nove corroboram com a classificação SARpy, como um valor experimental, dois com boa confiabilidade, cinco com confiabilidade moderada e apenas um com baixa confiabilidade. A maioria dos compostos, mais precisamente 29 compostos foram classificados em faixas próximas dos valores desses compostos. Quatro foram classificados com boa confiabilidade, 10 com moderado e 15 com baixa confiabilidade. Apenas oito compostos foram classificados em faixas não vizinhas de valores; por exemplo, Tóxico-1 e Tóxico-3, que foram previstos com boa confiabilidade e seis com baixa confiabilidade. Nenhuma discordância total entre os dois métodos de classificação foi verificada, ou seja, os mesmos compostos classificados como não-tóxico e tóxico-1 em ambos os métodos. A maioria dos compostos (34) foi classificada pelo CM como mais tóxica do que o método SARpy. A alta concordância dos valores previstos de pLC₅₀ das três abordagens ($r > 0,766$), CM, ToxRead e T.E.S.T., é visualizada na matriz de correlação (Tabela 10). Como visualizado na predição LC₅₀ de *D. magna*, CM e T.E.S.T. foram mais bem correlacionados ($r = 0,883$).

Tabela 8 - Os MPO potencialmente tóxicos selecionados contra *D. magna* e *P. promelas*

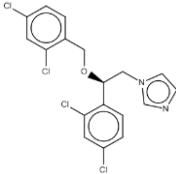
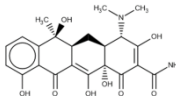
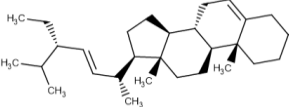
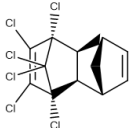
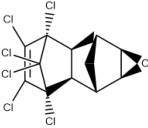
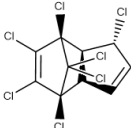
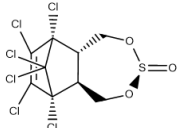
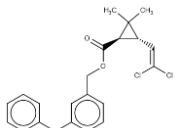
ID	Classe dos compostos	MPO	Estruturas
6	Antibiótico	Miconazole	
11		Tetracycline	
43	Hormonais	Stigmasterol	
49	Organoclorados	Aldrin	
51		Dieldrin	
52		Heptachlor	
53		Endosulfan	
55	Piretroide	Permethrin	

Tabela 9 - Predição qualitativa de CM e SARpy/IRF e confiabilidade da predição SARpy/IRF. A classificação é baseada nos valores de intervalo de LC₅₀. Tóxico 1, 2, 3 e não-tóxico, respectivamente, valores de LC₅₀ entre 1 e 10 mg / L, 10 e 50 mg / L, 50 e 100 mg / L, e > 100 mg / L. Em negrito estão destacados a combinação de previsão de ambas as abordagens

ID	CM	SARpy/IRF	Confiabilidade
1		NON-Tóxico	baixa
2		Tóxico-2	baixa
3		Tóxico-2	baixa
4		Tóxico-2	baixa
5		Tóxico-3	baixa
6	Tóxico-1	Tóxico-2	baixa
7		Tóxico-3	baixa
8		Tóxico-3	baixa
9	Tóxico-3	Tóxico-3	boa
10	Tóxico-3	Tóxico-3	baixa
11		Tóxico-3	baixa
12	Tóxico-2	Tóxico-3	moderada
13	Tóxico-2	Tóxico-3	baixa
14		NON-Tóxico	baixa
15	Tóxico-3	Tóxico-2	boa
16	Tóxico-2	Tóxico-3	boa
17	Tóxico-1	Tóxico-2	baixa
18	Tóxico-3	Tóxico-2	boa
19		Tóxico-3	baixa
20	Tóxico-2	Tóxico-2	boa
21	Tóxico-2	Tóxico-3	baixa
22	Tóxico-2	Tóxico-3	moderada
23		Tóxico-3	baixa
24		Tóxico-2	baixa
25	Tóxico-2	Tóxico-2	moderada

26		Tóxico-2	baixa
27	Tóxico-2	Tóxico-3	baixa
28	Tóxico-2	Tóxico-3	moderada
29		Tóxico-2	baixa
30		Tóxico-3	baixa
31		Tóxico-2	baixa
32	Tóxico-3	NON-Tóxico	moderada
33		Tóxico-2	baixa
34	Tóxico-3	NON-Tóxico	moderada
35	Tóxico-2	Tóxico-2	moderada
36	Tóxico-1	Tóxico-2	baixa
37	Tóxico-1	Tóxico-3	baixa
38	Tóxico-2	Tóxico-2	moderada
39	Tóxico-2	Tóxico-2	moderada
40	Tóxico-1	Tóxico-2	moderada
41	Tóxico-1	Tóxico-2	moderada
42	Tóxico-1	Tóxico-2	moderada
43	Tóxico-1	Tóxico-2	baixa
44	Tóxico-1	Tóxico-2	moderada
45	Tóxico-2	Tóxico-3	baixa
46	Tóxico-1	Tóxico-2	baixa
47	Tóxico-2	Tóxico-2	moderada
48	Tóxico-3	Tóxico-2	baixa
49	Tóxico-1	Tóxico-3	baixa
50	Tóxico-1	Tóxico-3	boa
51	Tóxico-1	Tóxico-2	baixa
52	Tóxico-1	Tóxico-3	baixa
53	Tóxico-1	Tóxico-3	baixa
54	Tóxico-1	Tóxico-3	baixa
55	Tóxico-1	Tóxico-1	experimental

56	Tóxico-3	Tóxico-1	boa
57	Tóxico-2	Tóxico-1	boa
58	Tóxico-1	Tóxico-2	moderada
59	Tóxico-2	NON-Tóxico	baixa
60	Tóxico-3	NON-Tóxico	baixa
61	Tóxico-3	NON-Tóxico	baixa
62		Tóxico-3	baixa
63	Tóxico-2	Tóxico-3	baixa
64	Tóxico-2	Tóxico-3	baixa

Tabela 10 - Matriz de correlação (r) dos valores preditos de LC₅₀ de *D. magna* pelos três métodos utilizados neste estudo para 64 MPO identificados em diferentes matrizes de água no Brasil. a) CM previu valores de LC₅₀ para 46 compostos, b) ToxRead para 45 e c) T.E.S.T. para 57 compostos

Modelos	CM^a	ToxRead^b	T.E.S.T.^c
RFCM	1	0.766	0.883
ToxRead		1	0.847
T.E.S.T.			1

Para *P. promelas*, dos 64 MPO identificados em diferentes fontes de água no Brasil apenas sete compostos obtiveram dados experimentais de LC₅₀, portanto o que não é possível verificar satisfatoriamente a predição dos métodos (Tabela 11). O ToxRead, por exemplo, previu apenas três compostos com valores experimentais. Todas as abordagens apresentaram valores mais baixos do coeficiente de predição $R_{ext}^2 < 0,39$ e valores mais altos de RMSE > 1,03 (Tabela 12).

Tabela 11 - Valores experimentais e previstos de pLC₅₀ de *P. promelas* obtidos pelas três diferentes abordagens

Id	CM	ToxRead	T.E.S.T.	Experimental
1	-	2.36	3.23	-
2	-	-	5.21	-
3	-	-	5.46	-
4	-	-	5.69	-
5	-	4.27	4.47	-
6	6.77	6.14	6.82	-
7	-	4.82	5.15	-
8	-	-	5.61	-
9	4.11	3.68	4.15	-
10	4.09	2.80	-	-
11	-	-	5.69	-
12	4.96	4.29	-	-
13	5.00	5.04	4.6	-
14	-	2.45	-	-
15	3.40	4.01	3.35	-
16	4.87	4.77	5.28	-
17	5.86	5.64	5.84	-
18	4.29	4.17	4.91	-
19	-	-	5.91	-
20	4.55	4.54	4.64	-
21	5.34	4.59	5.4	-
22	4.80	4.09	3.68	-
23	-	5.69	-	-
24	-	8.63	-	-
25	4.87	4.33	4.13	-
26	-	-	-	-
27	5.32	5.34	5.77	-
28	4.71	5.35	4.75	4.54

29	--	-	4.69	-
30	-	-	5.94	-
31	-	-	6.48	-
32	4.01	-	4.88	-
33	-	-	6.85	-
34	3.77	-	4.1	-
35	5.41	5.71	5.62	-
36	6.16	5.72	5.95	-
37	5.79	4.80	5.15	-
38	5.43	5.06	5.58	-
39	5.32	5.37	5.31	-
40	5.62	5.53	5.57	-
41	5.58	5.20	5.74	-
42	5.80	5.53	5.31	-
43	6.30	5.75	6.49	-
44	5.53	5.06	5.37	-
45	5.07	-	5.71	-
46	6.24	-	5.62	-
47	5.00	4.93	5.05	-
48	4.04	4.29	-	-
49	7.33	-	6.31	7.09
50	6.26	4.20	5.8	6.61
51	7.26	-	7.34	7.21
52	7.28	-	6.37	6.75
53	5.64	-	6.09	8.48
54	5.69	4.84	6.36	-
55	7.47	7.39	7.67	7.21
56	3.95	5.38	4.93	-
57	5.05	5.07	5.75	-
58	6.05	5.38	5.73	-
59	4.98	3.37	4.81	-

60	3.71	2.88	3.46	-
61	3.47	2.69	3.19	-
62		6.69	6.69	-
63	4.89	5.41	4.82	-
64	5.32	5.26	4.88	-

Tabela 12 - Coeficiente de determinação e o RMSE dos valores estimados de LC₅₀ de *D. magna* de 20 compostos de MPOs identificados em diferentes fontes de água no Brasil

Modelos	CM	ToxRead	T.E.S.T.
R _{ex} ^{t2}	0.22	0.14	0.38
RMSE	1.11	1.47	1.03

Entre as 64 MPO, todos os modelos prevêem vários compostos 6, 43, 49, 51, 52 e 55 (Tabela 8 e Tabela 11) com valores mais elevados de pLC₅₀ > 6,30. Os compostos 49, 51, 52 e 55 também foram preditos como mais tóxicos para *D. magna*. A toxicidade do composto 53 (Tabela 8 e Tabela 11) foi subestimada (pLC₅₀ < 6,09) muito inferior ao seu pLC₅₀ experimental = 8,48, para ambas as abordagens, CM e T.E.S.T.

4 CONCLUSÃO

Conseguimos desenvolver modelos aplicáveis de forma confiável para prever a toxicidade aquática de poluentes químicos identificados em diferentes fontes de água no Brasil contra *D. magna* e *P. promelas*. Os modelos também mostraram desempenho superior ou similar quando comparados com várias outras abordagens populares, como VEGA, T.E.S.T. e CAESAR (DEMETRA, EPA, SARpy e toxRead). Os resultados obtidos para *D. magna* mostraram um menor valor de validação de previsão, no entanto, uma precisão maior em relação aos valores previstos para 20 MPO. Os modelos desenvolvidos neste estudo podem ser utilizados como ferramentas de triagem para a predição de toxicidade aquática de compostos orgânicos contra ambos os organismos.

REFERÊNCIAS

- AN, X. et al. Treating organic cyanide-containing groundwater by immobilization of a nitrile-degrading bacterium with a biofilm-forming bacterium using fluidized bed reactors. **Environmental Pollution**, v. 237, p. 908–916, jun. 2018.
- BARBOSA, M. O. et al. Occurrence and removal of organic micropollutants: An overview of the watch list of EU Decision 2015/495. **Water Research**, v. 94, p. 257–279, 2016.
- BARREIROS, L. et al. Analysis of 17- β -estradiol and 17- α -ethinylestradiol in biological and environmental matrices - A review. **Microchemical Journal**, v. 126, p. 243–262, 2016.
- BIELLEN, A. et al. Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. *Water Research*, v. 126, p. 79–87, 2017.
- BREIMAN, L. E. O. Random Forests. *Machine Learning*, v. 45, p. 5–32, 2001.
- BUI, X. T. et al. Multicriteria assessment of advanced treatment technologies for micropollutants removal at large-scale applications. **Science of the Total Environment**, v. 563–564, p. 1050–1067, 2016.
- CAPUZZI, S. J. et al. Computer-Aided Discovery and Characterization of Novel Ebola Virus Inhibitors. **Journal of Medicinal Chemistry**, v. 61, n. 8, p. 3582–3594, 2018.
- CCANCCAPA, A. et al. Pesticides in the Ebro River basin: Occurrence and risk assessment. **Environmental Pollution**, v. 211, n. April, p. 414–424, 2016.
- CHERKASOV, A. et al. QSAR modeling: Where have you been? Where are you going to? **Journal of Medicinal Chemistry**, v. 57, n. 12, p. 4977–5010, 2014.
- CONSONNI, V.; BALLABIO, D.; TODESCHINI, R. Comments on the definition of the Q2 parameter for QSAR validation. **Journal of Chemical Information and Modeling**, v. 49, n. 7, p. 1669–1678, 2009.
- CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 3, p. 682–692, 2002.
- CRUCIANI, G.; PASTOR, M.; GUBA, W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. **European Journal of Pharmaceutical Sciences**, v. 11, n. SUPPL. 2, 2000.
- DE MORAIS E SILVA, L. et al. Predictive ecotoxicity of MoA 1 of organic chemicals using in silico approaches. **Ecotoxicology and Environmental Safety**, v. 153, n. September 2017, p. 151–159, 2018.

DEARDEN, J. C.; ROTUREAU, P.; FAYET, G. QSPR prediction of physico-chemical properties for REACH. **SAR and QSAR in Environmental Research**, v. 24, n. 4, p. 279–318, 2013.

DEEB, A. A. et al. Suspect screening of micropollutants and their transformation products in advanced wastewater treatment. **Science of the Total Environment**, v. 601–602, p. 1247–1253, 2017.

DESBIOLLES, F. et al. Occurrence and ecotoxicological assessment of pharmaceuticals: Is there a risk for the Mediterranean aquatic environment? **Science of the Total Environment**, v. 639, p. 1334–1348, 2018.

DOS PASSOS MENEZES, P. et al. Docking and physico-chemical properties of α - and β -cyclodextrin complex containing isopulegol: a comparative study. **Journal of Inclusion Phenomena and Macrocyclic Chemistry**, v. 85, n. 3–4, p. 341–354, 2016.

EBELE, A. J.; ABOU-ELWAFI ABDALLAH, M.; HARRAD, S. Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment. **Emerging Contaminants**, v. 3, n. 1, p. 1–16, mar. 2017.

FEDEILA, M. et al. Biodegradation of anionic surfactants by *Alcaligenes faecalis*, *Enterobacter cloacae* and *Serratia marcescens* strains isolated from industrial wastewater. **Ecotoxicology and Environmental Safety**, v. 163, n. August, p. 629–635, 2018.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of Chemical Information and Modeling**, v. 50, n. 7, p. 1189–1204, 2010.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Curation of chemogenomics data. **Nature Chemical Biology**, v. 11, n. 8, p. 535, 2015.

GASTEIGER, J. Chemoinformatics: Achievements and challenges, a personal view. **Molecules**, v. 21, n. 2, 2016.

GAVRILESCU, M. et al. Emerging pollutants in the environment: Present and future challenges in biomonitoring, ecological risks and bioremediation. **New Biotechnology**, v. 32, n. 1, p. 147–156, 2015.

GOLBRAIKH, A. et al. Data set modelability by QSAR. **Journal of Chemical Information and Modeling**, v. 54, n. 1, p. 1–4, 2014.

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

GRAMATICA, P.; CASSANI, S.; SANGION, A. Aquatic ecotoxicity of personal care products: QSAR models and ranking for prioritization and safer alternatives' design. **Green Chemistry**, v. 18, n. 16, p. 4393–4406, 2016.

GRAMATICA, P.; PAPA, E.; SANGION, A. QSAR modeling of cumulative environmental end-points for the prioritization of hazardous chemicals. **Environmental Science: Processes and Impacts**, v. 20, n. 1, p. 38–47, 2018.

GROS, M. et al. Screening and prioritization of micropollutants in wastewaters from on-site sewage treatment facilities. **Journal of Hazardous Materials**, v. 328, p. 37–45, 2017.

HAMID, H.; LI, L. Y.; GRACE, J. R. Review of the fate and transformation of per- and polyfluoroalkyl substances (PFASs) in landfills. **Environmental Pollution**, v. 235, p. 74–84, 2018.

HEYS, K. A. et al. Risk assessment of environmental mixture effects. **RSC Advances**, v. 6, n. 53, p. 47844–47857, 2016.

KOBAYASHI, T. et al. Estrogen alters gonadal soma-derived factor (GsdF)/Foxl2 expression levels in the testes associated with testis-ova differentiation in adult medaka, *Oryzias latipes*. **Aquatic Toxicology**, v. 191, n. May, p. 209–218, 2017.

LOCATELLI, M. A. F.; SODRÉ, F. F.; JARDIM, W. F. Determination of Antibiotics in Brazilian Surface Waters Using Liquid Chromatography–Electrospray Tandem Mass Spectrometry. **Archives of Environmental Contamination and Toxicology**, v. 60, n. 3, p. 385–393, 10 abr. 2011.

LUO, Y. et al. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. **Science of the Total Environment**, v. 473–474, p. 619–641, 2014.

MACHADO, K. C. et al. A preliminary nationwide survey of the presence of emerging contaminants in drinking and source waters in Brazil. **Science of The Total Environment**, v. 572, p. 138–146, dez. 2016.

MANDARIC, L. et al. Contamination sources and distribution patterns of pharmaceuticals and personal care products in Alpine rivers strongly affected by tourism. **Science of the Total Environment**, v. 590–591, p. 484–494, 2017.

MANSANO, A. S. et al. Effects of diuron and carbofuran pesticides in their pure and commercial forms on *Paramecium caudatum*: The use of protozoan in ecotoxicology. **Environmental Pollution**, v. 213, p. 160–172, 2016.

MELO, S. M.; BRITO, N. M. Analysis and Occurrence of Endocrine Disruptors in Brazilian Water by HPLC-Fluorescence Detection. **Water, Air, & Soil Pollution**, v. 225, n. 1, p. 1783, 6 jan. 2014.

MUNZ, N. A. et al. Pesticides drive risk of micropollutants in wastewater-impacted streams during low flow conditions. **Water Research**, v. 110, p. 366–377, 2017.

NEALE, P. A. et al. Integrating chemical analysis and bioanalysis to evaluate the contribution of wastewater effluent on the micropollutant burden in small streams. **Science of the Total Environment**, v. 576, p. 785–795, 2017.

NICOLAS, C. I. et al. Rapid experimental measurements of physicochemical properties to inform models and testing. **Science of the Total Environment**, v. 636, p. 901–909, 2018.

NÖDLER, K.; VOUTSA, D.; LICHA, T. Polar organic micropollutants in the coastal environment of different marine systems. **Marine Pollution Bulletin**, v. 85, n. 1, p. 50–59, 2014.

PETRIE, B.; BARDEN, R.; KASPRZYK-HORDERN, B. A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. **Water Research**, v. 72, n. 0, p. 3–27, 2015.

RIVERA-UTRILLA, J. et al. Pharmaceuticals as emerging contaminants and their removal from water. A review. **Chemosphere**, v. 93, n. 7, p. 1268–1287, 2013.

ROY, P. P.; ROY, K. On some aspects of variable selection for partial least squares regression models. **QSAR and Combinatorial Science**, v. 27, n. 3, p. 302–313, 2008.

SANTHOSH, C. et al. Role of nanomaterials in water treatment applications: A review. **Chemical Engineering Journal**, v. 306, p. 1116–1137, 2016.

SODRÉ, F. F. et al. Assessing selected estrogens and xenoestrogens in Brazilian surface waters by liquid chromatography–tandem mass spectrometry. **Microchemical Journal**, v. 96, n. 1, p. 92–98, set. 2010.

SUN, J. et al. Organic contamination and remediation in the agricultural soils of China: A critical review. **Science of the Total Environment**, v. 615, p. 724–740, 2018.

TAVAKOL, M. et al. Determining multivariate analysis sampling frequency for monitoring contamination caused by trout farms. **Polish Journal of Environmental Studies**, v. 26, n. 1, p. 337–346, 2017.

TODESCHINI, R. **Molecular Descriptors for Chemoinformatics**, Volumes I & II. [s.l.: s.n.].

TOOLARAM, A. P. et al. Initial hazard screening for genotoxicity of photo-transformation products of ciprofloxacin by applying a combination of experimental and in-silico testing. **Environmental Pollution**, v. 211, p. 148–156, 2016.

TROPSHA, A.; GOLBRAIKH, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. **Current pharmaceutical design**, v. 13, n. 34, p. 3494–504, jan. 2007.

TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR and Combinatorial Science**, v. 22, n. 1, p. 69–77, 2003.

WANG, Q.; WANG, P.; YANG, Q. Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. **Science of the Total Environment**, v. 621, p. 990–999, 2018.

YANG, Y. et al. Occurrences and removal of pharmaceuticals and personal care products (PPCPs) in drinking water and water/sewage treatment plants: A review. **Science of the Total Environment**, v. 596–597, p. 303–320, 2017.

YOU, L. et al. Investigation of pharmaceuticals, personal care products and endocrine disrupting chemicals in a tropical urban catchment and the influence of environmental factors. **Science of The Total Environment**, v. 536, p. 955–963, dez. 2015.

ZAKHAROV, A. V. et al. QSAR Modeling and Prediction of Drug-Drug Interactions. **Molecular Pharmaceutics**, v. 13, n. 2, p. 545–556, 2016

CAPÍTULO IV

CAPÍTULO IV - AVALIAÇÃO DA SEGURANÇA QUÍMICA DE PRODUTOS DE TRANSFORMAÇÃO DE LIXIVIADO DE ATERRO SANITÁRIO FORMADO DURANTE O PROCESSO FENTON ³

RESUMO

Compostos químicos orgânicos identificados no lixiviado de aterro bruto (LL) localizado na região metropolitana de Campina Grande, no nordeste do Brasil e seus produtos de transformação (TPs) formados durante o tratamento de Fenton foram estudados sob as diretrizes do REACH no contexto da avaliação de segurança química. 197 estruturas químicas únicas foram elucidadas, dentro delas, 154 compostos identificados no LL bruto e 82 compostos foram detectados no LL tratado, representando 39 compostos persistentes (PCs) e 43 TPs. Modelos *in silico* foram desenvolvidos para identificar e priorizar o potencial perigo/risco desses compostos para o meio ambiente e a para saúde humana. Os modelos revelaram que a biodegradabilidade dos TPs foi melhorada pelos processos de Fenton e um ligeiro aumento dos efeitos ecotoxicológicos observados do LL tratado em relação aos compostos presentes no LL bruto. Nenhuma diferença significativa foi encontrada para receptores de hidrocarboneto de arila (AhR), elementos de resposta antioxidante (ARE) e mutagenicidade através do teste de Ames. Comportamento semelhante de amostras LL brutas e tratadas foi observado para biodegradabilidade, mutagenicidade em *Tetrahymena pyriformis*, *Daphnia magna*, *Pimephales promelas*, ARE, AhR e AMES. No geral, nossos resultados indicaram que as amostras de LL bruto e tratado apresentaram perfis de atividades semelhantes para todos os parâmetros.

Palavras-chave: Processo de oxidação avançado, processo Fenton, Produtos de transformação, Perigos de produtos químicos, Quiminformática.

³ Capítulo submetido no Journal of Hazards Materials

ABSTRACT

Organic chemicals identified in raw landfill leachate (LL) located in the metropolitan region of Campina Grande, in the northeast of Brazil and their transformation products (TPs) formed during the Fenton treatment were studied under REACH guidelines in the context of the chemical safety assessment. 197 unique chemical structures were elucidated, within them, 154 compounds were identified in raw LL and 82 compounds were detected in the treated LL, representing 39 persistent compounds (PCs) and 43 TPs. *In silico* models were developed to identify and prioritize the potential hazard/risk of these compounds to the environment and society. The models revealed that the biodegradability of TPs was improved by the Fenton processes and a slight increase of the ecotoxicological effects observed of the LL treated compared to the compounds present in raw LL. No significant differences were found for aryl hydrocarbon receptor (AhR), antioxidant response element (ARE), and Ames mutagenicity. Similar behavior of both raw and treated LL samples was observed for biodegradability, *T. pyriformis*, *D. magna*, *P. promelas*, ARE, AhR, and AMES mutagenicity. Overall, our results indicated that raw and treated LL samples have similar activity profiles for all *endpoints* except biodegradability.

Keywords: Advanced oxidation process, Fenton process, Transformation products, Hazards of chemicals, Cheminformatics.

1 INTRODUÇÃO

O desenvolvimento econômico, a industrialização e a urbanização levaram à enorme geração de resíduos sólidos e perigosos, o que é um problema ambiental significativo tanto nos países em desenvolvimento quanto nos desenvolvidos. Globalmente, o aterro sanitário ainda é o método preferido para a disposição de resíduos sólidos devido ao seu baixo custo de operação (AZIZ; AMR, 2016). Um estudo recente mostrou que a produção anual de resíduos sólidos urbanos em todo o mundo atingiu 2,01 bilhões de toneladas e espera-se chegar a 3,40 bilhões de toneladas até 2050 (KAZA et al., 2018). No entanto, uma das desvantagens mais significativas dos métodos de disposição em aterro está relacionada à geração de lixiviado, um líquido altamente poluído formado pela decomposição de resíduos orgânicos e água que se infiltra e percola através do material residual (KJELDSEN et al., 2002). O efluente lixiviado é uma água residual com baixa biodegradabilidade, alta toxicidade, alta concentração de orgânicos refratários, contaminantes químicos inorgânicos, nitrogênio amoniacal, metais pesados, fenóis, pesticidas e materiais orgânicos xenobióticos (KJELDSEN et al., 2002; ROBINSON, 2017; SLACK; GRONOW; VOULVOULIS, 2005). O lixiviado pode e geralmente inclui um conjunto altamente preocupante e diversos de contaminantes, como substâncias per- e polifluoroalquil (HAMID; LI; GRACE, 2018), retardadores de chama bromados (GALLEN et al., 2016), substâncias de extinção de ultravioleta (DENG et al., 2018; ISKANDER et al., 2018), compostos farmacologicamente ativos, antibióticos e produtos de cuidados pessoais (CLARKE et al., 2015; YI et al., 2017). Consequentemente, se o lixiviado atingir o solo, as águas superficiais ou subterrâneas, ou se não for tratado e cuidadosamente descartado no meio ambiente, esse tipo de água residual pode causar vários efeitos nocivos à vida selvagem e à saúde humana (BARRIOS RESTREPO et al., 2017; GHOSH; GUPTA; THAKUR, 2015; NIKA et al., 2020).

O lixiviado é coletado e frequentemente transportado para tratamento de águas residuais. No entanto, os tratamentos secundários (por exemplo, processo convencional de lodo ativado) e terciários (como filtração e desinfecção) se mostraram insuficientes para a remoção de vários poluentes orgânicos (CAPODICI et al., 2019; RENOUE et al., 2008; RIZZO et al., 2019a, 2019b). Processos de oxidação avançados (AOPs), que envolvem a formação de espécies altamente oxidantes e não seletivas, como radicais hidroxila (HO^*),

têm demonstrado desempenhar um papel importante no tratamento de águas residuais e são capazes de degradar diferentes tipos de poluentes orgânicos (GAUTAM; KUMAR; LOKHANDWALA, 2019; GORITO et al., 2017). Dentre os diferentes POA, o tratamento Fenton ($\text{Fe}^{2+} + \text{H}_2\text{O}_2$) tem sido amplamente estudado e é considerado uma tecnologia promissora para a remoção de poluentes orgânicos em lixiviados de aterro (DENG; ENGLEHARDT, 2006; NAKHATE; PATIL; MARATHE, 2018; WANG et al., 2016).

Embora esses métodos sejam considerados eficazes para a remoção de poluentes orgânicos, em alguns casos, a geração dos produtos de transformação (TPs) durante os processos de tratamento avançados ainda é uma preocupação. Esses TPs podem preservar o modo de ação dos compostos parentais e podem ser mais tóxicos que seus precursores, sendo, portanto, uma preocupação para o homem e o meio ambiente (ESCHER; FENNER, 2011). Infelizmente, muitas vezes, os TPs não foram identificados nem regulamentados (ZAHN et al., 2019). Além disso, a quantidade e diversidade de TPs impedem uma avaliação de risco abrangente, principalmente devido à complexidade de sua análise experimental, que geralmente é demorada e cara (KHALEEL et al., 2019; KÜMMERER et al., 2019). Portanto, a avaliação experimental não é adequada para a triagem sob demanda da ampla variedade de compostos TPs, o que dificulta o conhecimento sobre os efeitos desses compostos em efluentes complexos.

Enquanto isso, as abordagens computacionais têm sido amplamente utilizadas para a avaliação e priorização de produtos químicos de interesse ambiental e humano (ALVES et al., 2018a; DE MORAIS E SILVA et al., 2018, 2019; KAR et al., 2020; KHAN; ROY; BENFENATI, 2019; MELNIKOV et al., 2016; TOROPOV et al., 2020) e se tornou uma solução prática para a tomada de decisão regulatória moderna (GRISONI et al., 2016; LI et al., 2017; LUO et al., 2019). Nos últimos anos, a legislação química europeia promoveu a gestão e a exploração de grandes volumes de dados para reduzir os testes em animais e economizar tempo e custos para melhorar as margens de segurança de novas substâncias químicas (GRAMATICA; CASSANI; SANGION, 2016; LILIENBLUM et al., 2008; MURATOV et al., 2020; NENDZA et al., 2018).

O objetivo principal deste estudo foi avaliar o potencial perigo de compostos químicos identificados no lixiviado de aterro sanitário bruto e seus TPs formados pelo processo Fenton. Para atingir este objetivo, nós (i) coletamos amostras brutas de chorume de

um aterro sanitário localizado na região metropolitana de Campina Grande, no nordeste do Brasil; (ii) realizou o tratamento Fenton do lixiviado do aterro; (iii) estruturas químicas identificadas por espectroscopia de massa; (iv) dados recolhidos, com curadoria e integrados sobre parâmetros de interesse como parte dos requisitos iniciais das diretrizes REACH no contexto da avaliação de segurança química; (v) desenvolveu modelos QSAR e MuDRA para esses terminais; e (vi) modelos desenvolvidos usados para criar um perfil abrangente de toxicidade/risco de produtos químicos e TPs investigados.

2 MATERIAL E MÉTODOS

2.1 Análise Experimental

2.1.1 Amostras de lixiviado e processo Fenton

As amostras brutas de lixiviado utilizadas neste estudo foram coletadas em um aterro sanitário localizado na região metropolitana de Campina Grande, Paraíba, região nordeste do Brasil. Os procedimentos experimentais de tratamento Fenton de lixiviado de aterro utilizado neste trabalho foram totalmente descritos por Dantas et al. (DANTAS et al., 2019).

2.1.2 Métodos analíticos para a elucidação de estruturas químicas

Os compostos brutos do lixiviado do aterro sanitário e seus produtos de transformação foram extraídos por sistema de extração líquido / líquido e identificados por cromatografia gasosa / espectrometria de massa (GC / MS QP - LCQ Fleet, Thermo Scientific) com uma coluna capilar Agilent HP-5 MS (30m x 0,25 mm x 0,25 mm, Agilent Technologies (EZ-Guard Columns) e uma fonte de ionização por electrospray. O volume injetado foi de 1,0 μL no modo splitless usando hélio como gás de arraste em uma coluna com fluxo de 1,0 $\text{mL}\cdot\text{min}^{-1}$. A temperatura inicial foi 40 ° C (mantido por 2 min), 10 ° $\text{C}\cdot\text{min}^{-1}$ a 70 ° C, em seguida, aquecimento a 5 ° $\text{C}\cdot\text{min}^{-1}$ a 250 ° C (mantido por 4 min). A fonte de íons MS foi operada a 200 ° C. Os espectros de massa foram obtidos em varredura completa no modo de íon positivo em proporções massa-carga (m/z) entre 50-650 amu. Os dados coletados foram identificados

pelo banco de dados do software Thermo Xcalibur (versão 2.2) que forneceu uma lista de possíveis produtos químicos fórmulas baseadas em m/ze distribuição isotópica. As estruturas dos subprodutos foram propostas de acordo com suas espectro de massa, fórmulas moleculares possíveis e tempo de retenção.

2.2 Abordagem quiminformatica

2.2.1 Conjunto de dados

Os conjuntos de dados foram cuidadosamente selecionados usando os fluxos de trabalho propostos por nosso grupo anteriormente (FOURCHES; MURATOV; TROPSHA, 2010, 2015, 2016). Resumidamente, quimiotipos específicos, como grupos aromáticos e nitro, bem como ligações duplas, foram normalizados e as configurações estereo foram removidas usando o Padronizador ChemAxon (v.19.27.0, ChemAxon, Budapeste, Hungria, <http://www.chemaxon.com>). Substâncias com subestrutura química indefinida e substâncias com peso molecular acima de 1.000 DA foram removidos. Contraíons, sais inorgânicos, compostos organometálicos e misturas foram removidos. Após a padronização estrutural, as duplicatas foram identificadas com o software HiT QSAR (KUZ'MIN; ARTEMENKO; MURATOV, 2008; TROPSHA, 2010) e cuidadosamente analisadas.

2.2.2 Descritores moleculares

Os modelos QSAR foram desenvolvidos usando descritores RDKit e impressões digitais Morgan (semelhantes a ECFP6), que foram calculados usando o software de quiminformática RDKit (<http://www.rdkit.org>, (RINIKER; LANDRUM, 2013)) executado no Python v.3.7 (<https://www.python.org>). As impressões digitais de Morgan foram geradas com raio 3 e vetor de bits de 2.048 bits. Os modelos MuDRA foram desenvolvidos usando impressões digitais Morgan (raio = 3 e 2048 bits), descritores de molécula inteira RDKit, chaves MACCS e impressões digitais Avalon.

2.2.3 Geração dos modelos de QSAR

Neste estudo, desenvolvemos modelos QSAR (CHERKASOV et al., 2014) e MuDRA (ALVES et al., 2018b) para prever a toxicidade potencial e o potencial de risco de compostos identificados a partir do lixiviado de aterro sanitário bruto e seus PTs formados durante o processo Fenton descrito na seção 2.1.1. Aqui, desenvolvemos modelos para oito *endpoints* de interesse, incluindo dois testes de propriedade físico-química: $\log K_{ow}$ (HILL, 2002) e biodegradabilidade estimada sob condições anaeróbicas, para um período de 28 dias (OECD 301D) (GHATTAS et al., 2017; Test No. 301: Ready Biodegradability, 1992); três ensaios para ecotoxicidade aquática: *Tetrahymena pyriformis* considerando a concentração causando inibição de crescimento de 50% ($pIGC_{50}$) após 48 horas de exposição [54], *Daphnia magna* com dados de concentração que mata 50% dos organismos (CL_{50}) no tempo de exposição 48 horas (Test No. 202: *Daphnia* sp. Acute Immobilisation Test, 2004) e *Pimephales promelas* usando dados de concentração letal média (CL_{50}) ao longo de uma duração de teste de 96 horas (Test No. 203: Fish, Acute Toxicity Test, 2019); teste de Ames determinado pela exposição de cepas de *Salmonella typhimurium* (MORTELMANS; ZEIGER, 2000; OECD, 2020) e dois ensaios para hepatotoxicidade em *Homo sapiens*, como receptor de hidrocarboneto de arila (AhR) contra linha celular HepG2-Ahr-luc (VELDHOEN et al., 2008) e elemento de resposta antioxidante (ARE) contra linha celular ARE-bla HepG2 (RAGHUNATH et al., 2018).

Modelos QSAR desenvolvidos em Python usando o algoritmo Random Forest (BREIMAN, 2001) implementado no scikit-learn (<https://scikit-learn.org/stable/>). Vinte rodadas de y-randomização foram realizadas para garantir que a previsibilidade não fosse dada ao acaso. Avaliamos o domínio de aplicabilidade (AD) dos modelos como $D_{cutoff} = \langle D \rangle + Zs$, onde Z é um parâmetro de limite de similaridade definido por um usuário (0,5 neste estudo), e $\langle D \rangle$ e s são a média e o padrão desvio, respectivamente, de todas as distâncias Euclidianas no espaço descritor multidimensional entre cada composto e seus vizinhos mais próximos para todos os compostos no conjunto de treinamento (GOLBRAIKH et al., 2003; TROPSHA; GOLBRAIKH, 2007). Os modelos MuDRA foram desenvolvidos em KNIME de acordo com o método descrito por Alves et al. (ALVES et al., 2018b) usando impressões

digitais Morgan (raio = 3 e 2048 bits), descritores de molécula inteira RDKit, chaves MACCS e impressões digitais Avalon.

2.2.4 Análise de agrupamento químico

A análise de agrupamento químico foi gerada pelo método *Sequential Agglomerative Hierarchical Non-overlapping* escrito em Python 3.7 (<https://www.python.org/>) e integrado em um fluxo de trabalho KNIME. Para esta análise, foram utilizados descritores RDKit de cada composto original. O agrupamento hierárquico foi realizado usando o pacote SciPy (<https://www.scipy.org/>) com base na distância euclidiana e o método de Ward (WARD, 1963). Resumidamente, o algoritmo gera um dendrograma e um mapa de calor da matriz de proximidade colorida de acordo com a similaridade/dissimilaridade química entre os compostos.

2.2.5 Análise estatística

As seguintes métricas estatísticas foram usadas para avaliar diferentes aspectos do desempenho dos modelos de classificação (Equações 1-5):

$$\begin{aligned} \text{Taxa de classificação correta (TCC)} & \qquad \qquad \qquad \text{Equação (1)} \\ & = \frac{\text{sensibilidade} + \text{especificidade}}{2} \end{aligned}$$

$$\text{Sensibilidade} = \frac{N_{\text{verdadeiros positivos}}}{N_{\text{verdadeiros positivos}} + N_{\text{falsos negativos}}} \qquad \text{Equação (2)}$$

$$\text{Especificidade} = \frac{N_{\text{verdadeiros negativos}}}{N_{\text{verdadeiros negativos}} + N_{\text{falsos positivos}}} \qquad \text{Equação (3)}$$

Valor de preditividade positiva (VPP)

$$\text{VPP} = \frac{N_{\text{verdadeiros positivos}}}{N_{\text{verdadeiros positivos}} + N_{\text{falsos positivos}}} \qquad \text{Equação (4)}$$

Valor de preditividade negativa (VPN)

$$VPN = \frac{N_{\text{verdadeiros negativos}}}{N_{\text{verdadeiros negativos}} + N_{\text{falsos negativos}}} \quad \text{Equação (5)}$$

3 RESULTADOS E DISCUSSÕES

3.1 Identificação de compostos em lixiviado de aterro sanitário e seus TPs formados durante o processo Fenton

A identificação dos compostos químicos presentes no lixiviado bruto do aterro sanitário e seus TPs formados durante o processo Fenton foi realizada através de um método analítico utilizando cromatografia em fase gasosa acoplada a um sistema de espectroscopia de massas CG/EM. Foram identificadas 197 estruturas químicas únicas, dentro delas, 154 compostos foram detectados em amostras de lixiviado sanitário bruto e 82 compostos foram detectados no lixiviado tratado, incluindo 39 compostos persistentes (CPs) que foram identificados antes e permaneceram inalterados após o tratamento. Além disso, foram identificados 43 produtos de transformação (PTs) formados pelo processo Fenton. Uma lista de compostos identificados pode ser encontrada em Materiais Suplementares (Tabela S1).

3.2 Análise de agrupamento hierárquico

Uma análise de agrupamento hierárquico foi realizada para identificar pequenos grupos de compostos com padrões de similaridade para o conjunto de dados heterogêneo de 197 compostos. Pela análise do mapa de calor e com base no dendrograma, um total de 17 e 13 agrupamentos foram identificados ao considerar uma distância euclidiana igual a 200 para definição de cluster para o lixiviado bruto e tratado, respectivamente (Figura 1). Entre os 17 grupos de lixiviados brutos, os grupos 4, 6, 8 e 14 registraram um número maior de compostos (Tabela 1). O cluster 4 agrupou 27 compostos (17,53%) caracterizados predominantemente por um anel monoaromático contendo frações alquil, como m-dimetil-

benzeno, fenilpropano, etilbenzeno, presença de grupos hidroxilas, aldeído, ácido carboxílico e moléculas alifáticas conectadas com heteroátomos contendo cadeias de C₄-C₁₀. O cluster 6 agrupou 11% dos compostos (n = 17). Este cluster consiste predominantemente em compostos aromáticos com um ou mais anéis fundidos a um anel de cinco membros ligado a grupos cloro, nitro e sulfonamida, com cadeias contendo C₆-C₁₄. O maior cluster 8, com 18% dos compostos (n = 28), consistiu principalmente de moléculas alifáticas e aromáticas com grupos cetona, álcool, amina, ácidos e álcoois graxos de cadeia média, com 6 a 12 átomos de carbono. Os compostos do cluster 14 (11% dos compostos, n = 17) foi formado principalmente por hidrocarbonetos acíclicos de cadeia ramificada, contendo em geral 12 a 16 átomos de carbono.

Figura 1 - Análise de agrupamento de 154 compostos identificados em lixiviado bruto (A) e 82 compostos identificados em lixiviado tratado (B).

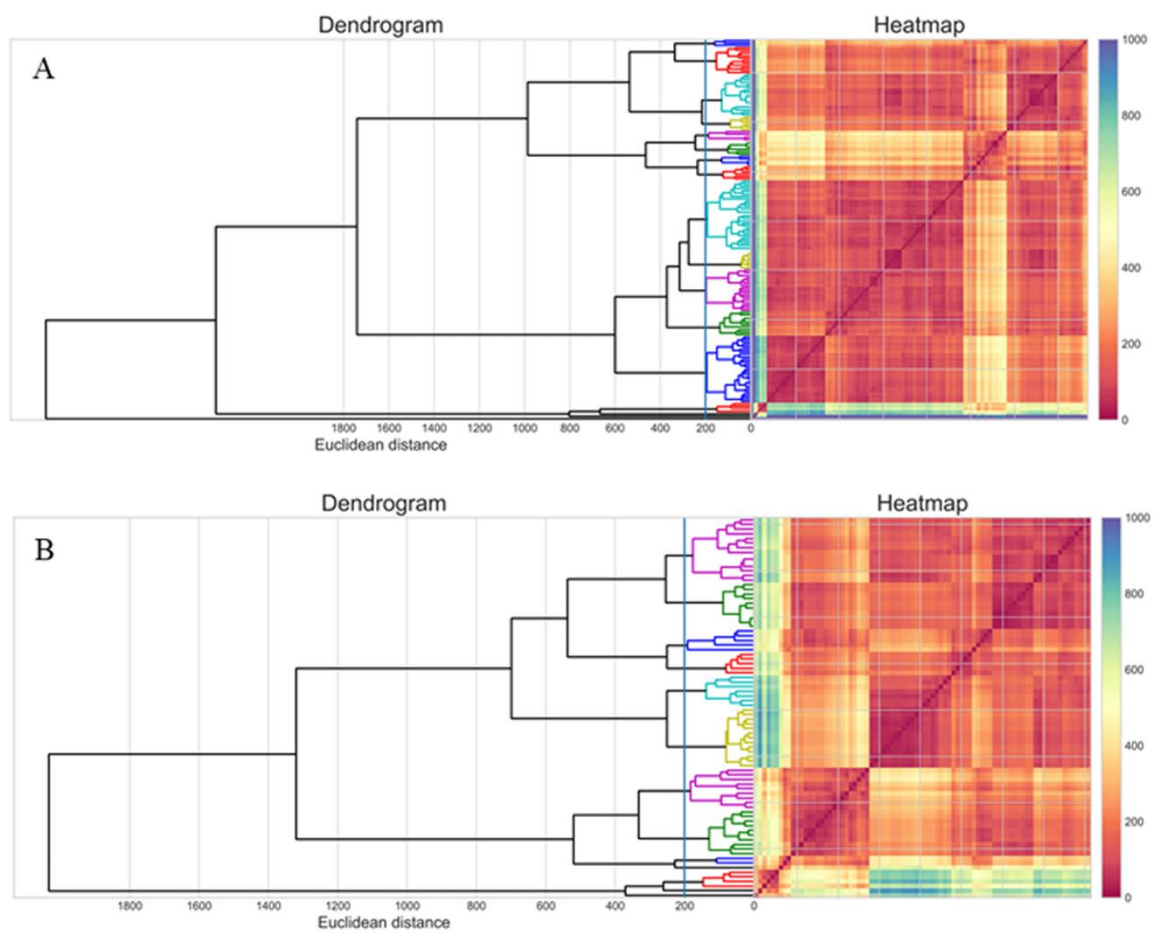
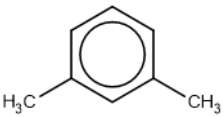
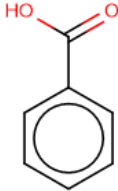
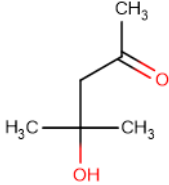
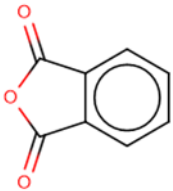

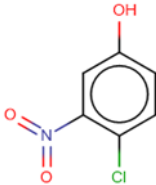
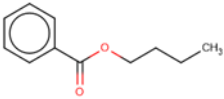
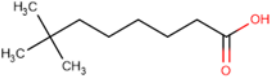
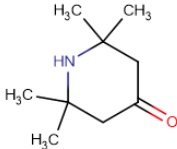
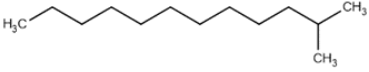
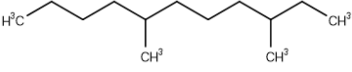
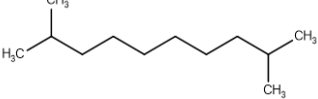


Tabela 1 - Exemplos de estruturas químicas presentes nos principais agrupamentos identificados para a amostra de lixiviado de aterro sanitário bruto

Cluster 4		
		
1,3-xylene	phenol	4-hydroxy-4-methylpentan-2-one
Cluster 6		
		
2-benzofuran-1,3-dione	1,1'-biphenyl	4-chloro-3-nitrophenol
Cluster 8		
		
butyl benzoate	7,7-dimethyloctanoic acid	2,2,6,6-tetramethylpiperidin-4-one
Cluster 14		
		
2-methyldodecane	3,7-dimethylundecane	2,9-dimethyldecane

No total, 13 agrupamentos foram detectados para os compostos presentes na amostra de lixiviado tratado por Fenton; os maiores clusters 6, 8, 12 e 13 foram analisados (Tabela 2). O cluster 6 contendo 10 compostos (12%) com 12 a 18 átomos de carbono, foi caracterizado por hidrocarbonetos de cadeias ramificadas e não ramificadas; e presença de

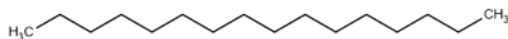
grupos hidroxila e carboxila de cadeia longa. A característica estrutural mais importante no cluster 8, que contém 13 compostos (16%), foi a presença de compostos monoaromático contendo grupos alquila, com 6 a 10 átomos de carbono. O cluster 12, contendo 12% dos compostos (n = 10), foi predominantemente formado por hidrocarbonetos saturados de cadeia ramificada, contendo 13 a 16 átomos de carbono. O cluster 13 (n =14, 17%) inclui alcanos ramificados e compostos de cadeia acíclica contendo grupos hidroxila e carboxila. Todos os outros clusters identificados são muito menores (incluindo menos de 10 compostos).

3.3 Modelos *in silico*

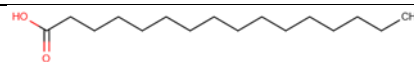
Os modelos demonstraram bom potencial preditivo para todos os *endpoints* de acordo com as melhores práticas de modelagem QSAR (CHERKASOV et al., 2014; TROPSHA, 2010). Os modelos log K_{ow} foram contínuos, apresentando Q^2 externo = 88%, erro quadrático médio = 0,41 e erro absoluto médio = 0,45. As principais características estatísticas dos modelos QSAR, MuDRA e consenso usados neste estudo são relatadas na Tabela 3.

Tabela 2 - Exemplos de estruturas químicas presentes nos principais agrupamentos identificados para a amostra de lixiviado tratado por Fenton

Cluster 6

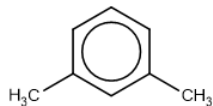


Hexadecane

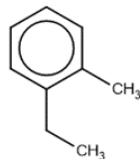


Palmitic acid

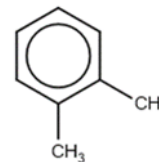
Cluster 8



1,3-xylene

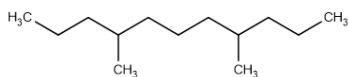


1-ethyl-2-methylbenzene

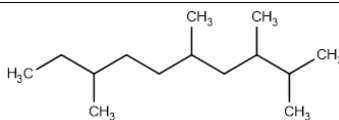


1,2-xylene

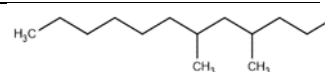
Cluster 12



4,8-Dimethylundecane

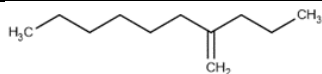


2,3,5,8-tetramethyldecane

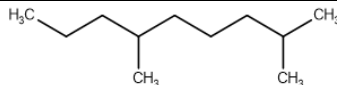


4,6-dimethyldodecane

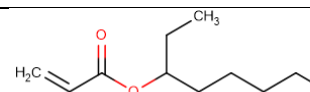
Cluster 13



4-methylidenedecane



2,6-dimethylnonane



nonan-3-yl prop-2-enoate

Tabela 3 - Características estatísticas dos modelos de QSAR, MuDRA e modelos de consenso para as predições das atividades selecionadas

Models	TCC	Sensibilidade	VPP	Especificidade	VPN	Coverage
Biodegradation (QSAR)	0.77	0.75	0.80	0.80	0.75	0.79
Biodegradation (MuDRA)	0.62	0.64	0.61	0.60	0.64	1.00
Consensus Biodegradation	0.70	0.70	0.71	0.70	0.70	0.90
Aquatic toxicity (<i>T. pyriformis</i>) (QSAR)	0.92	0.94	0.90	0.90	0.94	0.87
Aquatic toxicity (<i>T. pyriformis</i>) (MuDRA)	0.85	0.87	0.87	0.82	0.82	1.00
Consensus Aquatic toxicity (<i>T. pyriformis</i>)	0.89	0.91	0.89	0.86	0.88	0.94
Aquatic toxicity (<i>D. magna</i>) (QSAR)	0.80	0.77	0.80	0.86	0.80	0.81
Aquatic toxicity (<i>D. magna</i>) (MuDRA)	0.76	0.74	0.76	0.77	0.76	1.00
Consensus Aquatic toxicity (<i>D. magna</i>)	0.78	0.76	0.78	0.82	0.78	0.91
Aquatic toxicity (<i>P. promelas</i>) (QSAR)	0.81	0.79	0.83	0.83	0.79	0.84
Aquatic toxicity (<i>P. promelas</i>) (MuDRA)	0.70	0.67	0.71	0.74	0.70	1.00
Consensus Aquatic toxicity (<i>P. promelas</i>)	0.76	0.73	0.77	0.79	0.75	0.92
AhR (QSAR)	0.79	0.84	0.76	0.74	0.83	0.78
AhR (MuDRA)	0.71	0.71	0.71	0.72	0.72	1.00
Consensus AhR	0.75	0.78	0.74	0.73	0.78	0.89
ARE (QSAR)	0.79	0.83	0.76	0.75	0.82	0.72
ARE (MuDRA)	0.65	0.62	0.65	0.68	0.65	1.00
Consensus ARE	0.72	0.73	0.71	0.72	0.74	0.86

Ames mutagenicity (QSAR)	0.85	0.83	0.83	0.87	0.87	0.80
Ames mutagenicity (MuDRA)	0.79	0.81	0.81	0.77	0.76	1.00
Consensus Ames mutagenicity	0.82	0.82	0.82	0.82	0.82	0.90

O TCC dos modelos QSAR e MuDRA, estavam na faixa de 65-92% e 62-85%, respectivamente. A diferença entre sensibilidade e especificidade foi inferior a 0,1 para todos os modelos e, portanto, esperamos um número semelhante de previsões corretas para compostos tóxicos e não tóxicos. Todos os modelos MuDRA usados neste estudo exibiram 100% de cobertura de previsão e os modelos QSAR cobertura de previsão de 72-87%. Modelos de consenso (MC), obtidos a partir do cálculo dos valores médios dos dois modelos categóricos gerados pela combinação de QSAR e MuDRA, foram utilizados neste estudo para as previsões dos compostos (Tabela 3). O modelo de classificação foi atribuído como 1 (ativo), 0,5 (inconclusivo) e 0 (não ativo). Para a biodegradabilidade foi atribuído valores 1 para prontamente biodegradável, 0,5 para biodegradabilidade inconclusiva e 0 para não prontamente biodegradável, Tabela 4.

Tabela 4 - Resultados de QSAR, MuDRA e previsões de consenso para as atividades selecionadas

LL bruto							
	Biodegradabilidade	<i>T. pyriformis</i>	<i>D. magna</i>	<i>P. promelas</i>	AMES	AhR	ARE
Tóxico/Não-biodegradável	53	57	37	16	8	5	8
Inconclusivo	19	18	11	23	3	4	12
Não-tóxico/Biodegradável	28	25	53	62	92	92	81
LL bruto sem compostos persistentes							
	Biodegradabilidade	<i>T. pyriformis</i>	<i>D. magna</i>	<i>P. promelas</i>	AMES	AhR	ARE
Tóxico/Não-biodegradável	57	53	33	17	10	5	9
Inconclusivo	15	19	11	22	4	5	12
Não-tóxico/Biodegradável	28	28	56	61	86	90	79
Lixiviado tratado							
	Biodegradabilidade	<i>T. pyriformis</i>	<i>D. magna</i>	<i>P. promelas</i>	AMES	AhR	ARE
Tóxico/Não-biodegradável	40	63	41	15	4	5	6
Inconclusivo	30	21	17	28	1	2	15
Não-tóxico/Biodegradável	29	16	41	57	95	93	79
Produtos de transformação							
	Biodegradabilidade	<i>T. pyriformis</i>	<i>D. magna</i>	<i>P. promelas</i>	AMES	AhR	ARE
Tóxico/Não-biodegradável	40	58	35	19	5	7	7
Inconclusivo	30	26	26	30	2	5	19
Não-tóxico/Biodegradável	30	16	40	51	93	88	74
Compostos persistentes							
	Biodegradabilidade	<i>T. pyriformis</i>	<i>D. magna</i>	<i>P. promelas</i>	AMES	AhR	ARE
Tóxico/Não-biodegradável	41	69	49	8	3	3	5
Inconclusivo	31	15	8	26	0	0	10
Não-tóxico/Biodegradável	28	15	44	64	97	97	85

3.3.1 Log K_{ow}

O coeficiente de partição octanol-água ($\log K_{ow}$, também chamado de $\log P$) é um parâmetro físico-químico essencial usado para indicar a tendência de uma substância se distribuir no ambiente aquático, o que pode afetar seu destino ambiental (HILL, 2002). Os valores de $\log K_{ow}$ previstos para as amostras de lixiviado de aterro sanitário bruto e o lixiviado tratado são mostrados na Tabela S1. Além disso, foram utilizadas as médias \pm desvios padrão dos valores de $\log K_{ow}$. De acordo com os resultados, o $\log K_{ow}$ foi ligeiramente maior para os compostos de lixiviado tratados ($\log K_{ow}$ de $4,92 \pm 2,53$) do que os compostos detectados no lixiviado bruto ($\log K_{ow}$ de $3,92 \pm 2,51$), indicando que os compostos de lixiviado tratados são mais hidrofóbicos. Os compostos de CPs revelaram-se menos hidrofílicos do que os compostos de PTs, $\log K_{ow}$ de $5,22 \pm 2,57$ e $\log K_{ow}$ de $4,64 \pm 2,49$, respectivamente. A presença de ácidos graxos de cadeia longa e alcanos ramificados acíclicos com cadeia longa aumentou significativamente o $\log K_{ow}$. Por outro lado, o anel mono-aromático de benzeno contendo átomos de oxigênio, grupo amino, nitro ou sulfonamida e moléculas alifáticas oxigenadas, ácidos graxos, álcoois alifáticos ou éteres, com peso molecular inferior denotou efeito significativo para o valor $\log K_{ow}$ mais baixo (ver Figura 2).

3.3.2 Biodegradabilidade

A biodegradação é um processo natural de degradação de substâncias por meio de microrganismos. Este parâmetro é uma medida importante usada para avaliar o destino dos produtos químicos no meio ambiente e determinar se uma substância tem potencial para causar efeitos adversos no meio ambiente (GHATTAS et al., 2017). Em geral, quanto menor a biodegradabilidade ou não degradabilidade de um composto, maior a probabilidade de persistência no ambiente [54]. A previsão de biodegradabilidade foi classificada como prontamente biodegradável (1), biodegradabilidade inconclusiva (0,5) e não prontamente biodegradável (0). A classificação dos produtos químicos detectados em amostras brutas de lixiviado de aterro mostrou que 28% dos compostos ($n = 43$) foram previstos como prontamente biodegradáveis, 19% ($n = 29$) como biodegradáveis inconclusivos e 53% ($n =$

82) como não prontamente biodegradável. Por outro lado, as amostras tratadas por Fenton resultaram em 29% (n = 24) de biodegradável prontamente, 30% (n = 25) de biodegradável inconclusivo e 40% (n = 33) de não biodegradável prontamente. Nesse caso, não foi observada diferença entre os PTs e os CPs (Tabela 4). Os resultados das análises indicam que esses compostos em amostras tratadas tendem a ser mais biodegradáveis do que os compostos originais. Isso corresponde aos resultados de um estudo recente de biodegradabilidade de lixiviados pelo processo de Fenton [64]. Grupo carboxílico e seus derivados, cetonas, hidroxilas e cadeias alifáticas foram grupos funcionais observados para biodegradabilidade, enquanto os anéis de benzeno contendo grupos alquila nas posições 1 e 3, bem como alquilos ramificados (C12-C15) foram encontradas como interferem para a biodegradabilidade (ver Figura 3).

Figura 2 – Estrutura química e predições correspondentes para $\log K_{ow}$.

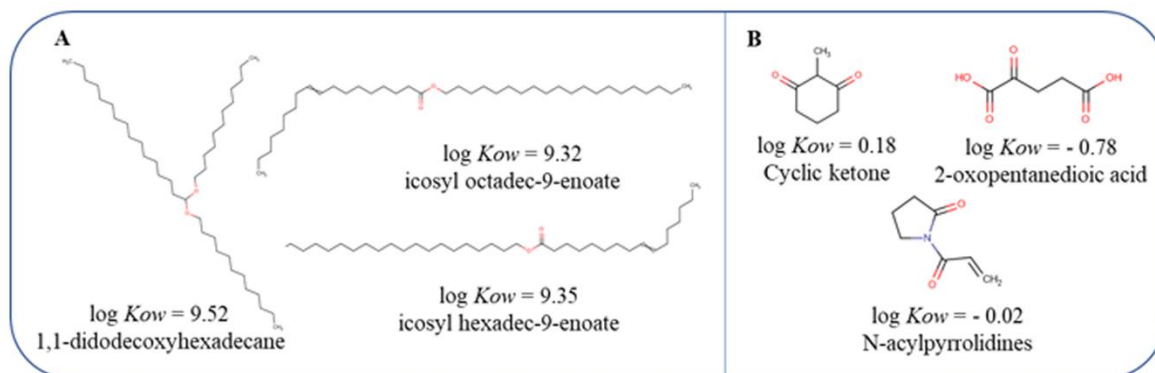
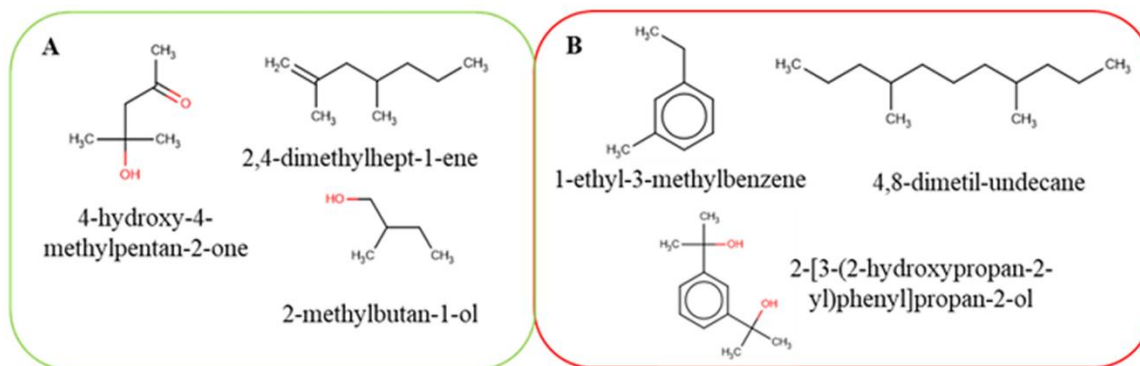


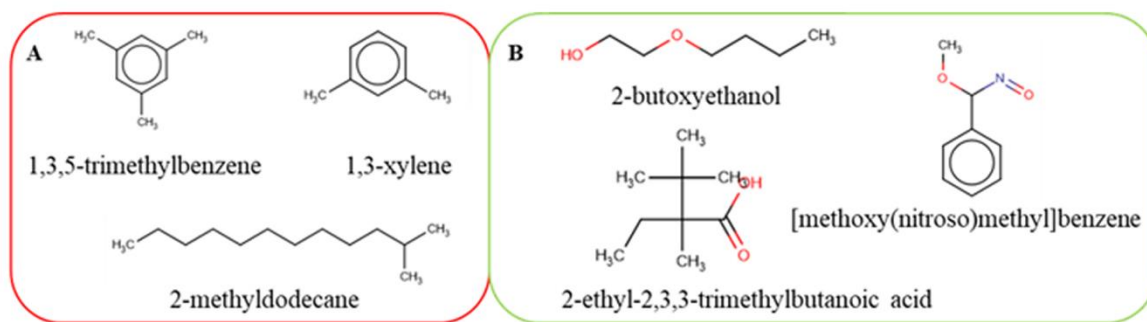
Figura 3 – Exemplo ilustrativo das estruturas químicas de compostos classificados como biodegradáveis (A) e não biodegradáveis (B).



3.3.3 *Tetrahymena pyriformis*

Tetrahymena pyriformis é um protozoário ciliado de vida livre encontrado em ecossistemas de água doce em todo o mundo, comumente usados como um importante bioindicador de poluição química no ambiente aquático (M. P. SAUVANT, D. PEPIN, 1999). Nossos resultados indicaram que 57% dos compostos em amostras de LL brutas (n = 88) foram considerados tóxicos, 18% (n = 28) inconclusivos e 25% não-tóxicos (n = 38). Entre os compostos em LL tratados, 63% (n = 52) e 21% (n = 17) foram preditos como tóxicos e inconclusivos, respectivamente, e 16% (n = 13) não tóxicos (Tabela 4). PTs e CPs mostraram perfis de atividades semelhantes. A presença de compostos monoaromático contendo substituinte metil e etil e hidrocarbonetos insaturados acíclicos ramificados ou não ramificados promoveu a toxicidade. Enquanto os hidrocarbonetos alifáticos com cadeia curta e anel monoaromático contendo grupos hidroxila, grupos amino, sulfonamida e heteroatomos foram evidenciados para os compostos preditos como não tóxicos (ver Figura 4).

Figura 4 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Tetrahymena pyriformis*.



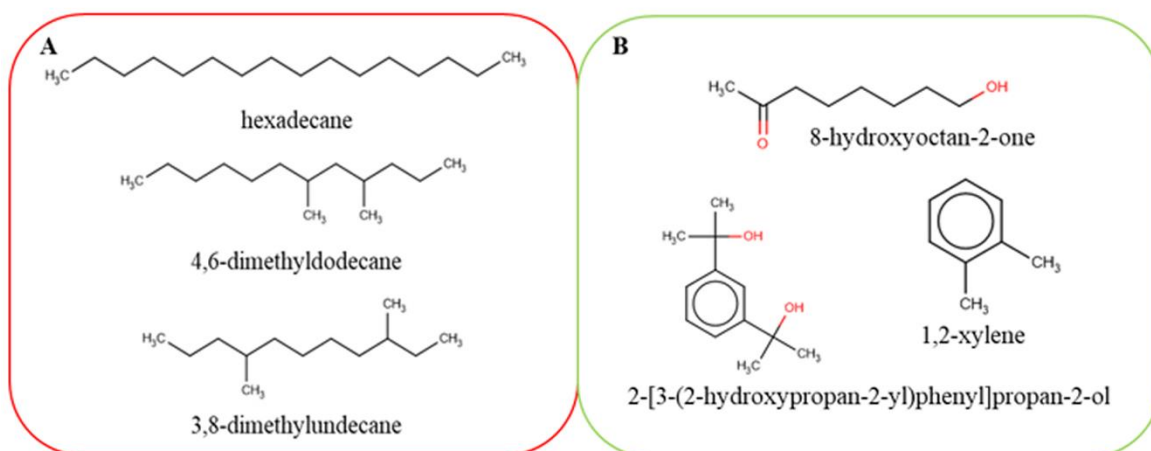
3.3.4 *Daphnia magna*

O crustáceo de água doce *D. magna* é considerado um organismo modelo nas avaliações de toxicidade ambiental de produtos químicos, uma vez que este organismo é muito sensível e desempenha um papel importante na cadeia alimentar (BARATA et al., 2017). A avaliação deste *endpoints* indicou que a maioria dos produtos químicos, 81 compostos (53%), em amostras de lixiviado de aterro bruto foram considerados não tóxicos

contra *D. magna*, com 37% (n = 57) tóxicos e 10% (n = 16) preditos como inconclusivo. O lixiviado tratado demonstrou maior proporção de compostos classificados como tóxicos (41%, n= 34) e menor proporção de não-tóxicos (41%, n = 34), em relação ao lixiviado bruto (ver Tabela 4). Resultados semelhantes foram obtidos por Dantas et al. [43]. Maior potencial de toxicidade bem com maior potencial de não-tóxicos foi observado para o CPs frente aos TPs, que, no entanto, apresentou mais compostos classificados como inconclusivos.

Hidrocarbonetos acíclicos ramificados ou não ramificados (C9-C26) foram principalmente tóxicos contra *D. magna*. As moléculas alifáticas oxigenadas contendo álcoois graxos de cadeia longa e derivados (> C17), cadeia de brometo de alquila, ácidos benzóicos com cadeia longa (C23-C25) e compostos monoaromáticos contendo um ou dois anéis de benzeno (C14-C17) com grupo hidroxila foram geralmente previstos como inconclusivos. Em geral, a presença de compostos com cadeias curtas (C4-C12) contendo derivados de benzila, grupos metil, etil e nitro, bem como compostos contendo funções oxigenadas foram comumente encontradas interferindo na toxicidade (ver Figura 5).

Figura 5 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Daphnia magna*.

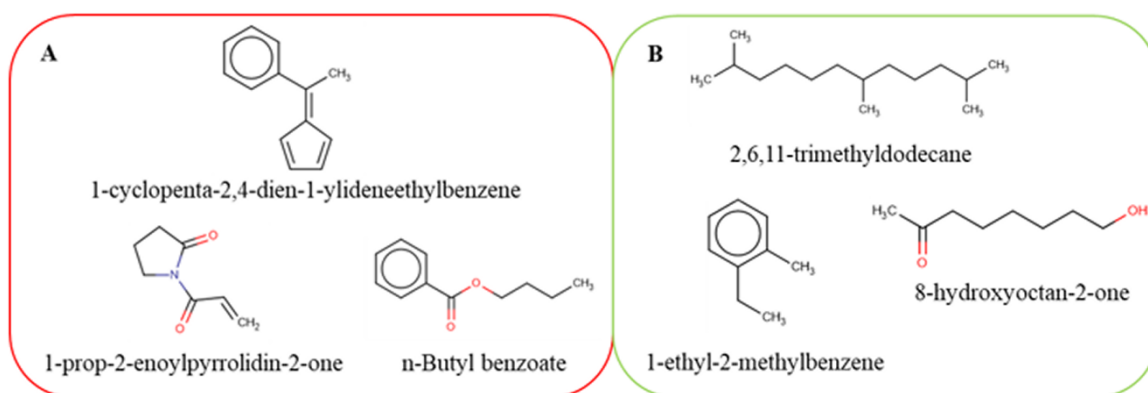


3.3.5 *Pimephales promelas*

Os peixes, e em particular o *Pimephales promelas*, desempenham um papel importante na avaliação da toxicidade devido ao seu papel importante nas cadeias

alimentares, sua tendência para bioacumular substâncias tóxicas, bem como sua sensibilidade a baixas concentrações de mutagênicos (Test No. 203: Fish, Acute Toxicity Test, 2019). Neste modelo, 16% dos compostos (n = 24) detectados em amostras de lixiviado de aterro foram previstos como tóxicos, 35 compostos foram previstos como inconclusivos (23%) e 95 (62%) não tóxicos. Por outro lado, 15% dos compostos (n = 12) presentes no LL tratado foram identificados como tóxicos, 28% (n = 23) como inconclusivos e 57% (n = 47) como não tóxicos. O que resultou em uma diminuição de 50% e 49% em termos de quantidade de compostos tóxicos e não tóxicos, respectivamente. Nesse caso, identificou-se os TPs como mais tóxicos que os PCs (Tabela 4). A toxicidade contra *P. promelas* foi influenciada principalmente pelo maior número de anéis e porções de ciclos na molécula contendo grupos funcionais como cetona, grupo carboxílico e aldeído. Ácidos graxos de cadeia longa (C12-C40), ésteres de ácidos graxos, ésteres de álcool graxo, a maioria deles saturados, e alcanos foram principalmente previstos como inconclusivos. Os compostos não tóxicos geralmente têm peso molecular inferior e consistem principalmente de alcanos ramificados, monoaromaticos ligados a halogênios, oxigênio, nitrogênio e enxofre; e compostos de cadeia alifática contendo grupos oxigenados como álcoois, ésteres, cetonas e carboxílicos (ver Figura 6).

Figura 6 - Exemplo ilustrativo das estruturas químicas de compostos classificados como tóxicos (A) e não tóxicos (B) para *Pimephales promelas*.



3.3.6 AhR e ARE

Receptor de hidrocarboneto de arila (AhR), um receptor nuclear presente na maioria dos tecidos de vertebrados, conhecido por mediar a toxicidade e os efeitos biológicos de compostos (SHAHIN et al., 2020) e o elemento de resposta antioxidante (ARE), uma parte da via de resposta ao estresse oxidativo associado a patogênese de uma variedade de doenças, desde câncer até neurodegeneração (RAGHUNATH et al., 2018), foram avaliados para melhorar a identificação de toxicidade química. Para ambos, encontramos um efeito previsto semelhante para amostras de lixiviado bruto e tratado. Mais de 90% e cerca de 80% dos compostos não apresentaram efeito positivo previsto nos ensaios AhR e ARE, respectivamente. Para esses *endpoints*, os PTs têm 10% mais compostos previstos como tóxicos do que os PCs (Tabela 4).

3.3.7 AMES

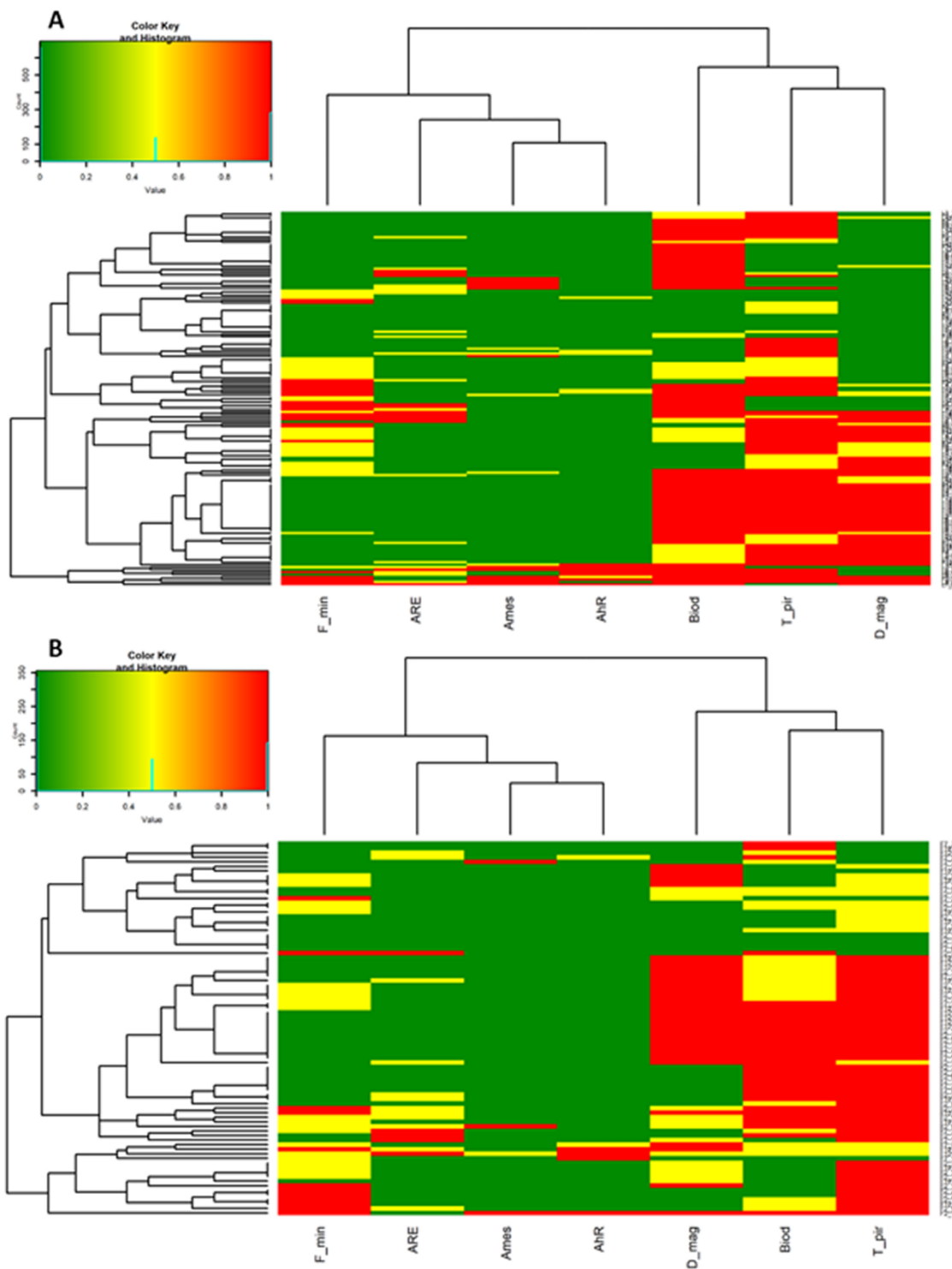
O teste de Ames, determinado pela exposição de cepas de *Salmonella typhimurium*, é o teste para avaliar o potencial efeito carcinogênico de produtos químicos mais utilizado em todo o mundo (OECD, 2020). Neste modelo, os resultados indicaram que em aterro de lixiviado bruto, apenas aromáticos mono e policíclicos (n = 17) contendo heteroátomos foram principalmente mutagênicos no teste de Ames. No aterro tratado, mais de 95% dos compostos foram preditos como não mutagênicos no teste de Ames (Tabela 4).

3.3.8 Análise de similaridade das atividades

Uma avaliação comparativa foi feita para avaliar as semelhanças/dissimilaridades dos desfechos investigados por meio de perfis de risco dos compostos do LL bruto e tratado. Para isso, utilizamos cluster de correlação por mapa de calor usando distância euclidiana (HEATMAP, [s.d.]). Conforme mostrado na Figura 7 e Figura. S1. Os resultados relatados no dendrograma, indicam que nossos sete pontos finais podem ser agrupados em dois grupos. O primeiro grupo indicou a correlação entre três desfechos, biodegradabilidade, *T. pyriformis* e *D. magna*. Por exemplo, *T. pyriformis* e *D. magna* foram agrupados, o que sugere que os

compostos de interesse irão realizar uma ação tóxica semelhante para ambos. O segundo grupo corresponde à correlação de pontos finais *P. promelas*, ARE, AMES e AhR, indicando que a correlação entre ARE e AMES é mais forte. Os resultados mostram que tanto as amostras de LL brutas quanto as tratadas têm um perfil de atividade semelhante para esses pontos finais.

Figura 7 - Mapa de calor para visualizar a matriz de correlação de diferentes *endpoints* de lixiviados bruto (A) e tratado (B). A cor verde está associada a compostos não tóxicos/biodegradáveis e vermelhos - a compostos tóxicos/não biodegradáveis.



4 CONCLUSÃO

Nossos resultados indicaram que o tratamento Fenton pode remover ~ 60% dos produtos químicos orgânicos encontrados no lixiviado do aterro. Com base no $\log K_{ow}$ estimado, os compostos lixiviados tratados mostraram uma menor solubilidade em água com uma fração maior de ácidos graxos de cadeia longa e alcanos ramificados acíclicos com uma cadeia longa. O tratamento Fenton também aumenta a proporção de produtos biodegradáveis de 47% para 60%. No entanto, os efeitos ecotoxicológicos observados do lixiviado de aterro sanitário tratado revelaram um ligeiro aumento em comparação com os compostos presentes no lixiviado de aterro bruto. *T. pyriformis* e *D. magna* foram considerados as espécies mais sensíveis, sugerindo que a avaliação de risco desses subprodutos em ambientes aquáticos não pode ser ignorada. Potencial mutagênico muito baixo do LL bruto e seus TPs foi detectado por meio de nossos modelos de Ames. O tratamento de Fenton não resulta em diferença significativa de potencial AhR e ARE entre LLs e TPs. Até onde sabemos, este é o primeiro estudo em que o produto de transformação durante a degradação do lixiviado sanitário pelo processo Fenton e seus perfis de risco foram estimados. Esperamos que nossos resultados forneçam suporte teórico para a avaliação da toxicidade do lixiviado de aterro sanitário e dos produtos de transformação de Fenton e a compreensão de seu potencial perigo.

REFERÊNCIAS

ALVES, V. M. et al. Chemical toxicity prediction for major classes of industrial chemicals: Is it possible to develop universal models covering cosmetics, drugs, and pesticides? **Food and Chemical Toxicology**, v. 112, p. 526–534, 2018a.

ALVES, V. M. et al. Multi-Descriptor Read Across (MuDRA): A Simple and Transparent Approach for Developing Accurate Quantitative Structure-Activity Relationship Models. **Journal of Chemical Information and Modeling**, v. 58, n. 6, p. 1214–1223, 2018b.

AZIZ, H. A.; AMR, S. A. **Control and Treatment of Landfill Leachate for Sanitary Waste Disposal**. [s.l: s.n.].

BARATA, C. et al. Validation of a two-generational reproduction test in *Daphnia magna*: An interlaboratory exercise. **Science of the Total Environment**, v. 579, p. 1073–1083, 2017.

BARRIOS RESTREPO, J. J. et al. Correlation between acute toxicity for *Daphnia magna*, *Aliivibrio fischeri* and physicochemical variables of the leachate produced in landfill simulator reactors. **Environmental Technology (United Kingdom)**, v. 38, n. 22, p. 2898–2906, 2017.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.

CAPODICCI, M. et al. Achievement of partial nitrification under different carbon-to-nitrogen ratio and ammonia loading rate for the co-treatment of landfill leachate with municipal wastewater. **Biochemical Engineering Journal**, v. 149, n. February, 2019.

CHERKASOV, A. et al. QSAR modeling: Where have you been? Where are you going to? **Journal of Medicinal Chemistry**, v. 57, n. 12, p. 4977–5010, 2014.

CLARKE, B. O. et al. Investigating landfill leachate as a source of trace organic pollutants. **Chemosphere**, v. 127, p. 269–275, 2015.

DANTAS, E. R. B. et al. Fenton treatment of sanitary landfill leachate: optimization of operational parameters, characterization of sludge and toxicology. **Environmental Technology (United Kingdom)**, v. 3330, 2019.

DE MORAIS E SILVA, L. et al. Predictive ecotoxicity of MoA 1 of organic chemicals using in silico approaches. **Ecotoxicology and Environmental Safety**, v. 153, n. September 2017, p. 151–159, 2018.

DE MORAIS E SILVA, L. et al. Predictive Computational Tools for Assessment of Ecotoxicological Activity of Organic Micropollutants in Various Water Sources in Brazil. **Molecular Informatics**, v. 38, n. 8–9, 2019.

DENG, Y. et al. Adsorption of UV-quenching substances (UVQS) from landfill leachate with activated carbon. **Chemical Engineering Journal**, v. 350, n. January, p. 739–746, 2018.

DENG, Y.; ENGLEHARDT, J. D. Treatment of landfill leachate by the Fenton process. **Water Research**, v. 40, n. 20, p. 3683–3694, 2006.

ESCHER, B. I.; FENNER, K. Recent advances in environmental risk assessment of transformation products. **Environmental Science and Technology**, v. 45, n. 9, p. 3835–3847, 2011.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of Chemical Information and Modeling**, v. 50, n. 7, p. 1189–1204, 2010.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Curation of chemogenomics data. **Nature Chemical Biology**, v. 11, n. 8, p. 535, 2015.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. **Journal of Chemical Information and Modeling**, v. 56, n. 7, p. 1243–1252, 2016.

GALLEN, C. et al. Occurrence and distribution of brominated flame retardants and perfluoroalkyl substances in Australian landfill leachate and biosolids. **Journal of Hazardous Materials**, v. 312, p. 55–64, 2016.

GAUTAM, P.; KUMAR, S.; LOKHANDWALA, S. Advanced oxidation processes for treatment of leachate from hazardous waste landfill: A critical review. **Journal of Cleaner Production**, v. 237, p. 117639, 2019.

GHATTAS, A. K. et al. Anaerobic biodegradation of (emerging) organic contaminants in the aquatic environment. **Water Research**, v. 116, p. 268–295, 2017.

GHOSH, P.; GUPTA, A.; THAKUR, I. S. Combined chemical and toxicological evaluation of leachate from municipal solid waste landfill sites of Delhi, India. **Environmental Science and Pollution Research**, v. 22, n. 12, p. 9148–9158, 2015.

GOLBRAIKH, A. et al. Rational selection of training and test sets for the development of

validated QSAR models. **Journal of Computer-Aided Molecular Design**, v. 17, n. 2–4, p. 241–253, 2003.

GORITO, A. M. et al. A review on the application of constructed wetlands for the removal of priority substances and contaminants of emerging concern listed in recently launched EU legislation. **Environmental Pollution**, v. 227, p. 428–443, 2017.

GRAMATICA, P.; CASSANI, S.; SANGION, A. Aquatic ecotoxicity of personal care products: QSAR models and ranking for prioritization and safer alternatives' design. **Green Chemistry**, v. 18, n. 16, p. 4393–4406, 2016.

GRISONI, F. et al. Expert QSAR system for predicting the bioconcentration factor under the REACH regulation. **Environmental Research**, v. 148, p. 507–512, 2016.

HAMID, H.; LI, L. Y.; GRACE, J. R. Review of the fate and transformation of per- and polyfluoroalkyl substances (PFASs) in landfills. **Environmental Pollution**, v. 235, p. 74–84, 2018.

HEATMAP. No Title. [s.d.].

HILL, E. F. **Wildlife toxicology of organophosphorus and carbamate pesticides**. [s.l.: s.n.].

ISKANDER, S. M. et al. A review of landfill leachate induced ultraviolet quenching substances: Sources, characteristics, and treatment. **Water Research**, v. 145, p. 297–311, 2018.

KAR, S. et al. **Ecotoxicological assessment of pharmaceuticals and personal care products using predictive toxicology approaches**. [s.l.: s.n.]. v. 22

KAZA, S. et al. **What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050**. [s.l.] The World Bank, 2018.

KHALEEL, N. D. H. et al. Studying the fate of the drug Chlorprothixene and its photo transformation products in the aquatic environment: Identification, assessment and priority setting by application of a combination of experiments and various in silico assessments. **Water Research**, v. 149, p. 467–476, 2019.

KHAN, K.; ROY, K.; BENFENATI, E. Ecotoxicological QSAR modeling of endocrine disruptor chemicals. **Journal of Hazardous Materials**, v. 369, n. January, p. 707–718, 2019.

KJELDSEN, P. et al. Present and long-term composition of MSW landfill leachate: A

review. **Critical Reviews in Environmental Science and Technology**, v. 32, n. 4, p. 297–336, 2002.

KÜMMERER, K. et al. Reducing aquatic micropollutants – Increasing the focus on input prevention and integrated emission management. **Science of the Total Environment**, v. 652, p. 836–850, 2019.

KUZ'MIN, V. E.; ARTEMENKO, A. G.; MURATOV, E. N. Hierarchical QSAR technology based on the Simplex representation of molecular structure. **Journal of computer-aided molecular design**, v. 22, n. 6–7, p. 403–21, jun. 2008.

LI, R. et al. Landfill Leachate Treatment by Electrocoagulation and Fiber Filtration. **Water Environment Research**, v. 89, n. 11, p. 2015–2020, 2017.

LILIENBLUM, W. et al. Alternative methods to safety studies in experimental animals: Role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH). **Archives of Toxicology**, v. 82, n. 4, p. 211–236, 2008.

LUO, X. et al. Rate constants of hydroxyl radicals reaction with different dissociation species of fluoroquinolones and sulfonamides: Combined experimental and QSAR studies. **Water Research**, v. 166, 2019.

M. P. SAUVANT, D. PEPIN, E. P. Tetrahymena Tetrahymena. **Science**, v. 38, n. 7, p. 1631–1669, 1999.

MELNIKOV, F. et al. Assessment of predictive models for estimating the acute aquatic toxicity of organic chemicals. **Green Chemistry**, v. 18, n. 16, p. 4432–4445, 2016.

MORTELMANS, K.; ZEIGER, E. The Ames Salmonella/microsome mutagenicity assay. **Mutation research**, v. 455, n. 1–2, p. 29–60, nov. 2000.

MURATOV, E. N. et al. QSAR without borders. **Chemical Society reviews**, v. 49, n. 11, p. 3525–3564, 2020.

NAKHATE, P. H.; PATIL, H. G.; MARATHE, K. V. Intensification of landfill leachate treatment by advanced Fenton process using classical and statistical approach. **Chemical Engineering and Processing - Process Intensification**, v. 133, n. October, p. 148–159, 2018.

NENDZA, M. et al. PBT assessment under REACH: Screening for low aquatic bioaccumulation with QSAR classifications based on physicochemical properties to replace BCF in vivo testing on fish. **Science of the Total Environment**, v. 616–617, p. 97–106,

2018.

NIKA, M. C. et al. Wide-scope target analysis of emerging contaminants in landfill leachates and risk assessment using Risk Quotient methodology. **Journal of Hazardous Materials**, v. 394, n. January, 2020.

OECD. Test Guideline No . 471 Bacterial Reverse Mutation Test. **Organisation for Economic Co-operation and Development guidelines**, n. 471, 2020.

RAGHUNATH, A. et al. Antioxidant response elements: Discovery, classes, regulation and potential applications. **Redox Biology**, v. 17, n. April, p. 297–314, 2018.

RENOU, S. et al. Landfill leachate treatment: Review and opportunity. **Journal of Hazardous Materials**, v. 150, n. 3, p. 468–493, 2008.

RINIKER, S.; LANDRUM, G. A. Open-source platform to benchmark fingerprints for ligand-based virtual screening. **Journal of Cheminformatics**, v. 5, n. 5, 2013.

RIZZO, L. et al. Consolidated vs new advanced treatment methods for the removal of contaminants of emerging concern from urban wastewater. **Science of the Total Environment**, v. 655, n. November 2018, p. 986–1008, 2019a.

RIZZO, L. et al. Tertiary treatment of urban wastewater by solar and UV-C driven advanced oxidation with peracetic acid: Effect on contaminants of emerging concern and antibiotic resistance. **Water Research**, v. 149, p. 272–281, 2019b.

ROBINSON, T. Removal of toxic metals during biological treatment of landfill leachates. **Waste Management**, v. 63, p. 299–309, 2017.

SHAHIN, N. N. et al. Potential role of aryl hydrocarbon receptor signaling in childhood obesity. **Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids**, v. 1865, n. 8, 2020.

SLACK, R. J.; GRONOW, J. R.; VOULVOULIS, N. Household hazardous waste in municipal landfills: Contaminants in leachate. **Science of the Total Environment**, v. 337, n. 1–3, p. 119–137, 2005.

Test No. 202: Daphnia sp. Acute Immobilisation Test. [s.l.] OECD, 2004.

Test No. 203: Fish, Acute Toxicity Test. [s.l.] OECD, 2019.

Test No. 301: Ready Biodegradability. [s.l.] OECD, 1992.

TOROPOV, A. A. et al. Use of the index of ideality of correlation to improve aquatic solubility model. **Journal of Molecular Graphics and Modelling**, v. 96, 2020.

- TROPSHA, A. Best practices for QSAR model development, validation, and exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, 2010.
- TROPSHA, A.; GOLBRAIKH, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. **Current pharmaceutical design**, v. 13, n. 34, p. 3494–504, jan. 2007.
- VELDHOEN, M. et al. The aryl hydrocarbon receptor links TH17-cell-mediated autoimmunity to environmental toxins. **Nature**, v. 453, n. 7191, p. 106–109, 2008.
- WANG, N. et al. A review on Fenton-like processes for organic wastewater treatment. **Journal of Environmental Chemical Engineering**, v. 4, n. 1, p. 762–787, 2016.
- WARD, J. H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, v. 58, n. 301, p. 236–244, 1963.
- YI, X. et al. Removal of selected PPCPs, EDCs, and antibiotic resistance genes in landfill leachate by a full-scale constructed wetlands system. **Water Research**, v. 121, p. 46–60, 2017.
- ZAHN, D. et al. Identification of potentially mobile and persistent transformation products of REACH-registered chemicals and their occurrence in surface waters. v. 150, p. 86–96, 2019.