



CENTRO UNIVERSITÁRIO DE ANÁPOLIS – UniEVANGÉLICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SOCIEDADE, TECNOLOGIA E MEIO
AMBIENTE

MODELOS COMPUTACIONAIS PARA AVALIAÇÃO PREDITIVA DE
ECOTOXICIDADE EM *Artemia salina*

BRUNO FRANCISCO CARDOSO DE LACERDA

ANÁPOLIS-GO

2019

BRUNO FRANCISCO CARDOSO DE LACERDA

**MODELOS COMPUTACIONAIS PARA AVALIAÇÃO PREDITIVA DE
ECOTOXICIDADE EM *Artemia salina***

Dissertação submetida ao Centro Universitário de Anápolis, como requisito parcial para obtenção do Título de Mestre em Ciências Ambientais. Área de concentração: Tecnologia e Meio Ambiente. Sociedade, Tecnologia e Meio Ambiente.

Orientador: Prof. Dr. Bruno Junior Neves.

Co-Orientador (a): Prof.^a Dr.^a Josana de Castro Peixoto

ANÁPOLIS-GO

2019

L131

Lacerda, Bruno Francisco Cardoso de.

Modelos computacionais para avaliação preditiva de ecotoxicidade em *Artemia Salina* / Bruno Francisco Cardoso de Lacerda – Anápolis: Centro Universitário de Anápolis – UniEvangélica, 2019. 47 p.; il.

Orientador: Prof. Dr. Bruno Junior Neves.

Co-Orientadora: Profa. Dra. Josana de Castro Peixoto

Dissertação (mestrado) – Programa de pós-graduação em Sociedade, Tecnologia e Meio Ambiente – Centro Universitário de Anápolis – UniEvangélica, 2019.

1. Toxicidade aquática 2. Modelagem preditiva 3. *Artemia Salina*
I. Neves, Bruno Junior II. Peixoto, Josana de Castro III. Título.
CDU 504

Catálogo na Fonte

Elaborado por Rosilene Monteiro da Silva CRB1/3038

BRUNO FRANCISCO CARDOSO DE LACERDA

**MODELOS COMPUTACIONAIS PARA AVALIAÇÃO PREDITIVA DE
ECOTOXICIDADE EM *Artemia salina***

Prof. Dr. Bruno Junior Neves
Centro Universitário de Anápolis, UniEvangélica
Presidente da Banca Examinadora – Orientador

Prof. Dr. Clarimar José Coelho
Centro Universitário de Anápolis, UniEvangélica
Membro da Banca Examinadora

Prof. Dr. Marcelo do Nascimento Gomes
Faculdade Metropolitana de Anápolis, FAMA
Membro da Banca Examinadora

DEDICATÓRIA

Dedico este trabalho primeiramente a Deus, pois tem sido o meu amparo e a força que me impulsiona a prosseguir. A minha formação profissional não poderia ter sido concretizada sem a ajuda de meus amáveis pais, companheiros e colegas Geraldo e Maria Suely, portadores de corações tão generosos que fizeram da arte de educar seus filhos uma missão de vida. Agradeço ainda minha esposa Rodryne eterna companheira, confidente, minha alma gêmea. Aos meus filhos Barbara e Rodrigo pelo simples fato de vocês estarem presente em minha vida. Amo vocês até onde minha alma alcança. Vocês são minha motivação e força para enfrentar todos e quaisquer desafios, sintam parte dessa nossa vitória.

AGRADECIMENTOS

Primeiramente a Deus e aos meus espíritos protetores, por me conceder sabedoria e paciência nas escolhas dos melhores caminhos. Por sempre me mostrarem que as provações vividas só nos dão mais força para enfrentarmos nossos desafios com mais serenidade.

Aos meus pais, Geraldo de Lacerda de Carvalho e Maria Suely Cardoso de Carvalho por terem me dado, educação, amor e por acreditarem em mim, nos meus sonhos, por terem sido compreensivos ao me incentivar a cada tropeço. Sem vocês seria impossível chegar até aqui. Obrigado por todo apoio, paciência, amor e colo

Aos meus irmãos Breno Cesar Cardoso de Lacerda e Bethânia de Lacerda Carvalho, que me apoiaram e compartilharam momentos de descontração e muita alegria e por sempre estarem presentes quando mais precisei.

À minha esposa, Rodryne Rodrigues de Carvalho Marciano Lacerda, companheira, amiga, confidente que me compreende, respeita e apoia. O seu amor e o simples fato da sua presença em minha vida são imprescindíveis. Obrigado por fazer parte da minha vida e caminhar ao meu lado.

Aos meus filhos, Barbara de Lacerda Carvalho e Bernardo de Lacerda Carvalho. Vocês foram e serão à motivação para o início, o andamento e o término de todo e qualquer soerguimento, por amá-los incondicionalmente. Só posso dizer muito obrigado e pedir lhes desculpas pelas ausências, mesmo quando estava presente. A vocês dedico não apenas este trabalho, mais a minha vida. Amo muito vocês!

Em especial, ao meu orientador Professor Dr. Bruno Junior Neves, obrigado pela oportunidade de crescimento pessoal e profissional, confiança, compreensão das minhas dificuldades e aflições, incentivo, paciência, dedicação, generosidade, carinho, disponibilidade, amizade, exemplo de tolerância, competência e humanismo em todos os momentos que busquei seu auxílio. Segue aqui minha admiração e respeito por acreditar e confiar em mim nesse trabalho.

À minha co-orientadora Prof. ^aDr.^a Josana de Castro Peixoto, pelo suporte disponibilizado, pelas correções e incentivos por ter encontrado potencial em mim, por ser minha companheira nessa caminhada, pelos conselhos amigos e pela sabedoria, na qual aprendi, cresci e amadureci.

Ao amigo e colaborador Josiel Araujo Lemes que além de sua dedicação, competência, sempre se mostrou tão solícito em minhas dúvidas, me ajudando sempre que no precisei. Você foi parte fundamental na realização deste sonho. Sem você não seria possível à execução deste trabalho.

Aos professores do Programa de Pós-graduação em Sociedade Tecnologia e Meio Ambiente – PPSTMA do Centro Universitário de Anápolis – UniEVANGÉLICA que com muito empenho, preocuparam-se em transmitir seus conhecimentos para formação de novos mestres.

Aos colegas do PPSTMA, que compartilharam comigo momentos difíceis, duvidosos, descontraídos, de aprendizado, crescimento pessoal e profissional até que chegássemos aqui.

“Aqui, no entanto, nós não olhamos para trás por muito tempo. Nós continuamos seguindo em frente, abrindo novas portas e fazendo coisas novas, porque somos curiosos e a curiosidade continua nos conduzindo por novos caminhos”

(Walt Disney)

RESUMO

O ambiente aquático é altamente complexo e diversificado, visto que compreende vários tipos de ecossistemas dentre os quais se encontram rios, lagos, estuários, mares e oceanos. Todos esses ecossistemas são produtos dinâmicos de interações complexas entre os componentes bióticos e abióticos característicos de cada um deles. A *Artemia salina* é um organismo de água salgada pertencente ao filo *Arthropoda*, disposto na família *Artemiidae*. Descrito como um crustáceo filtrador de ecossistemas talássicos, que se alimenta principalmente de bactérias, algas unicelulares, pequenos protozoários e detritos que são dissolvidos no meio. A intensificação do uso de agrotóxicos associado a expansão de áreas de produção agrícola tem ameaçado a biodiversidade do ecossistema aquático e seus serviços prestados. O presente trabalho objetivou a construção e validação de modelos de QSTR úteis para a predição de ecotoxicidade de compostos em *Artemia salina*. Metodologicamente, a modelagem dos parâmetros toxicológicos pode ser representada como um processo de três partes. Inicialmente, foi construído, preparado e padronizado o maior conjunto de dados de compostos com dados de LC₅₀ para *A. salina* disponível na literatura. Em seguida, todas as estruturas químicas de molécula foram codificadas em descritores moleculares. Posteriormente, métodos de aprendizado de máquina *Random Forest (RF)* e *Support Vector Machine (SVM)* foram utilizados para estabelecer relações quantitativas entre os descritores moleculares e propriedade toxicológica estudada. A preditividade do modelo QSTR obtido foi calculada utilizando métricas apropriadas, as quais avaliaram a habilidade do modelo em prever corretamente a toxicidade de compostos avaliados experimentalmente. Dos modelos construídos, o que se destacou em robustez e preditividade foi a combinação de FeatMorgan_2 + Random Forest (CCR = 0,82; SE = 0,80; SP = 0,85; PPV= 0,84; NPV= 0,81; k= 0,65; Cobertura= 0,66. Os modelos gerados serão implementados em servidores online para que os utilizadores possam realizar predições.

Palavras-chave: toxicidade aquática; modelagem preditiva; *Artemia salina*.

ABSTRACT

The aquatic environment is highly complex and diversified, since it comprises several types of ecosystems, including rivers, lakes, estuaries, seas and oceans. All these artemia components characteristic of each of them. *Artemia salina* is a saltwater organism belonging to the phylum Arthropoda, arranged in the family Artemiidae. Described as a crustacean filtering of thalassic ecosystems, it feeds mainly on bacteria, unicellular algae, small protozoa and debris that are dissolved in the medium. The intensification of the use of agrochemicals associated with the expansion of agricultural production areas has threatened the biodiversity of the aquatic ecosystem and its services. The present work aimed at the construction and validation of QSTR models for the prediction of ecotoxicity of compounds in *Artemia salina*. Methodologically, the modeling of toxicological parameters can be represented as a three-part process. Initially, the largest data set of compounds with LC₅₀ data for *A. salina* available in the literature was constructed, prepared and standardized. Then all chemical structures of molecules were converted into molecular descriptors. Subsequently, Random Forest (RF) and Support Vector Machine (SVM) machine learning methods were used to establish quantitative relationships between the molecular descriptors and the toxicological property studied. The predictivity of the obtained model was solculated using appropriate metrics, which evaluate the ability of the model to correctly predict the toxicity of experimentally evaluated compounds. From the models constructed, what stood out in robustness and predictability was the combination of FeatMorgan_2 + Random Forest (CCR = 0,82; SE = 0,80; SP = 0,85; PPV = 0,84; NPV = 0,81; k = 0.65 Coverage = 0.66. The generated models will be implemented in online servers so that users can make predictions.

Key-words: aquatic toxicity; predictive modeling; *Artemia salina*

LISTA DE FIGURAS

Figura 1. Áreas compreendidas pela toxicologia ambiental.....	Erro! Indicador não definido.
Figura 2. Ciclo de vida da <i>Artemia salina</i>	21
Figura 3. <i>Artemia salina</i>	23
Figura 4. Exemplos de inseticidas, fungicidas e herbicidas da biblioteca de dados.....	27
Figura 5. Principais etapas envolvidas na construção e validação de modelos de QSTR.....	28
Figura 6. Representação do plano de separação gerado por SVM (A) e do consenso das predições obtidas pelas árvores de decisão no método de Random Forest (B.	40
Figura 7. Alertas estruturais identificados a partir da interpretação mecanística do modelo FeatMorgan_2. Os fragmentos ou átomos coloridos em roxo possuem contribuição positiva para a toxicidade em <i>A. salina</i>	40

LISTA DE SIGLAS E ABREVIATURAS

$\mu\text{g/L}$ – Micrograma por Litro

$\mu\text{g/mL}$ – Micrograma por mililitro

$\mu\text{M/mL}$ – Micromolar por mililitro

ABNT - Agência Brasileira de Normas Técnicas

CCR – Taxa de Classificação Correta

CETESB - Companhia Ambiental do Estado de São Paulo

DA – Dominio de Aplicabilidade

DL₅₀ - Dose Letal para 50% da amostragem.

ECFP – Morgan

FCFP – FeatMorgan

g/L – Grama por Litro

K – Kappa

LC₅₀ - Concentração Letal para 50% da amostragem.

mg/mL – Miligrama por mililitro

mm – Milímetro

NPV – Valor Preditivo Negativo

OECD - Organização para a Cooperação e Desenvolvimento Económico.

PPSTMA - Programa de Pós-graduação em Sociedade Tecnologia e Meio Ambiente.

PPT - Partes Por Trilhão

PPV – Valor Preditivo Positivo

QSTR - Relações Quantitativas entre Estrutura Química e Toxicidade.

RF - *Random Forest*.

SE – Sensibilidade

SP - Especificidade

SVM - *Support Vector Machine*.

TAS – Toxicidade Frente a *Artemia salina*

LISTA DE QUADROS

Quadro 1. Classificação Científica da <i>Artemia Salina</i>	22
Quadro 2. Resultados Estatísticos dos Modelos de QSTR construídos para predição de Ecotoxicidade em <i>A. Salina</i>	36

SUMÁRIO

1. INTRODUÇÃO	14
2. OBJETIVOS	16
2.1 OBJETIVO GERAL	16
2.2 OBJETIVOS ESPECÍFICOS	16
3. REFERENCIAL TEÓRICO	17
3.1 ECOSISTEMAS AQUÁTICOS	17
3.1 ECOTOXICOLOGIA	17
3.2 <i>Artemia salina</i>	21
3.3 AVALIAÇÃO DE ECOTOXICIDADE	25
3.4 TOXICOLOGIA COMPUTACIONAL	27
3.5 DESCRITORES MOLECULARES	28
3.6 APRENDIZADO DE MÁQUINA	29
3.7 BOAS PRÁTICAS DE DESENVOLVIMENTO E VALIDAÇÃO EM QSTR	30
4. MATERIAIS E MÉTODOS	31
4.1 INTEGRAÇÃO, PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS ...	31
4.2 BALANCEAMENTO DO CONJUNTO DE DADOS	32
4.3 DESCRITORES MOLECULARES	32
4.4 CONTRUÇÃO DOS MODELOS	Erro! Indicador não definido.
4.5 VALIDAÇÃO CRUZADA EXTERNA DE 5-FOLDS (5FECV)	33
4.6 AVALIAÇÃO DE PREDITIVIDADE DOS MODELOS	33
Pr(e) = $\frac{VP + FP}{VP + FN + VN + FN + VN + FPN}$ (7)	34
4.7 DOMÍNIO DE APLICABILIDADE	34
4.8 INTERPRETAÇÃO MECANÍSTICA DOS MODELOS	35
5. RESULTADOS E DISCUSSÃO	36
5.1. PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS	36
5.1.1 PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS	36
5.1.2 MODELOS DE QSTR	37
5.1.3 ALERTAS ESTRUTURAIS	40
6. CONCLUSÕES E PERSPECTIVAS	41
7. REFERÊNCIAS BIBLIOGRÁFICAS	42

1. INTRODUÇÃO

Múltiplas fontes de contaminação, provenientes do uso prolongado de produtos químicos, ameaçam a integridade dos recursos hídricos, a nível mundial (POTT et al., 1989; DÍAZ-CRUZ; BARCELÓ, 2008). A consequência desse fenômeno é o risco de exposição da biota aquática a estes agentes tóxicos. A origem de tais agentes é variada, por exemplo, de resíduos tóxicos provenientes de efluentes industriais e drenagem agrícola (MONTANHA et al., 2011).

A toxicologia é a ciência que estuda os efeitos de produtos químicos ao nível do organismo individual, ou suas partes constituintes. Essa enfatiza as bases mecanística dos efeitos nocivos e as condições sob as quais eles ocorrem. Já a toxicologia aquática, se encarrega especificamente de estudar os efeitos de compostos químicos e outros xenobióticos em organismos aquáticos (SILVA; SANTOS, 2007).

De maneira geral, agentes toxicológicos podem causar variados danos a um sistema biológico, desde alterar uma função ou levando-o à morte, sob certas condições de exposição (SILVA; SANTOS, 2007). No caso específico de sistemas hídricos, análises ecotoxicológicas são aplicadas no monitoramento de águas e efluentes industriais (SOUSA, 2002).

Para se realizar a avaliação completa da toxicidade de uma grande quantidade de agentes tóxicos, em diferentes doses e concentrações, por meio de ensaios experimentais, dispende-se muito tempo e recursos. Além disso, resultaria em um problema ético, no que diz respeito ao uso de animais experimentais em tais testes. Com base nesta problemática, propõe o desenvolvimento de métodos alternativos de estudos que utilizam inteligência artificial. Um exemplo, são os modelos computacionais de Relações Quantitativas entre Estrutura Química e Toxicidade (QSTR), modelos que visam desenvolver relações quantitativas estatisticamente entre estrutura ou propriedades moleculares de compostos orgânicos abrangendo drogas, inseticidas, produtos químicos industriais e poluentes com seus dados de toxicidade com um objetivo de sua avaliação preditiva de risco e controle de risco ambiental (ROY, 2004).

Modelos de QSTR são capazes de analisar o potencial toxicológico e a identificação prévia de potenciais contaminantes de maneira mais rápida e com menor custo que os testes convencionais, que utilizam seres vivos. Esses avanços tecnológicos permitem construir, validar e implementar modelos computacionais de avaliação ecotoxicológica, ainda no

momento do desenvolvimento de um agente químico, por exemplo no desenvolvimento de candidatos a agrotóxicos (BRASIL, 1989; IBAMA, 1996; BRASIL, 2002).

2. OBJETIVOS

2.1 OBJETIVO GERAL

Construir modelos de QSTR robustos e preditivos para a avaliação de ecotoxicidade de compostos orgânicos em *Artemia salina*.

2.2 OBJETIVOS ESPECÍFICOS

- Compilar, integrar e preparar um conjunto de compostos com dados de toxicidade aguda em *Artemia salina*;
- Construir e validar modelos de QSTR binários através da combinação de diferentes tipos de descritores moleculares com os métodos de aprendizado de máquina *Floresta Randômica* (RF, do inglês *Random Forest*) e *Máquina de Suporte de Vetores* (SVM, do inglês *Support Vector Machine*);
- Realizar a interpretação mecanística dos modelos de QSTR construídos;

3. REFERENCIAL TEÓRICO

3.1 ECOSSISTEMAS AQUÁTICOS

A expansão das áreas de produção agrícola e intensificação do uso de agrotóxicos têm ameaçado a biodiversidade e os serviços ecossistêmicos prestados (TYLIANAKIS, 2013). Estes produtos podem entrar na cadeia solo-água-planta, representando uma perigosa fonte direta e indireta de contaminação para os seres vivos.

Os ecossistemas aquáticos podem ser divididos em: ecossistemas de água salgada e ecossistemas de água doce. Os ecossistemas de água salgada são também chamados de “talássicos” e correspondem aos mares e oceanos (RAMOS; AZEVEDO, 2010). O grande número e variedade dos ambientes, leníticos ou lântico e lóticos, de modo geral facilita estabelecer certos padrões de considerável validade regional, que permitem agrupar os corpos d’água conforme algumas características limnológicas.

Para identificar os efeitos destas substâncias sobre a biota aquática, tem sido utilizados, nestas últimas décadas, testes de toxicidade com organismos de águas continentais, estuarinas e marinhas, em condições laboratoriais e/ou de campo (BERTOLETTI, 1990). Tais testes possibilitam estabelecer limites permissíveis para substâncias químicas e avaliar o impacto de poluentes sobre os organismos aquáticos dos corpos receptores (CONAMA, 2005).

3.1 ECOTOXICOLOGIA

A ecotoxicologia surgiu na década de 60, sendo uma ciência que integra os princípios de várias ciências como a Ecologia, Toxicologia, Limnologia, Estatística, Química, Oceanografia, dentre outras. Esta ciência permite estudar e avaliar com precisão os efeitos de agentes físicos e químicos (substâncias naturais ou sintéticas) sobre os organismos vivos, populações e comunidades em seus ecossistemas, incluindo as formas de transporte dessas substâncias e suas interações com o ambiente (ZAGATTO, 2015; COSTA, 2008)

A história mostra que em diferentes épocas da existência, estudiosos vêm investigando as respostas de diferentes organismos vivos aos variados tipos de intervenção para se conhecer as características do meio em que se está inserido (MAGALHÃES, FERRÃO FILHO, 2008).

No âmbito do gerenciamento, pode ser conceituada como uma ciência que fornece dados aos gestores ambientais, quanto ao monitoramento ambiental e previsibilidade de impacto da poluição para diferentes sistemas (ZAGATTO, 2015).

Os testes de toxicidade são ferramentas que fornecem dados para avaliação de efeitos tóxicos nos sistemas biológicos e dimensionar toxicidade relativa, para prever ação em grande escala, em seres humanos e grandes ecossistemas. Consistem em ensaios laboratoriais, controlados, como objetivo de estimar a toxicidade de substâncias, nos quais os organismos-testes são expostos a diferentes concentrações de amostra e os efeitos são observados e quantificados (COSTA et al., 2008).

Protocolos padronizados de testes toxicológicos nos Estados Unidos da América (EUA) e na Europa, foram sendo desenvolvidos a partir da década de 80, em resposta à preocupação de se reduzir custos de produção e minimizar os efeitos de poluentes nos seres humanos, sem com isso retroceder na produção de bens de consumo. Nos EUA este movimento ficou conhecido como “*Clean Water Act*”, projeto pelo qual a Agência Americana de Proteção Ambiental pôde realizar atividades de vigilância em efluentes com atividades industriais (USEPA, 2002).

A preocupação com a qualidade aquática, gira em torno do pressuposto de que, os sistemas aquáticos são importantes portadores de agentes contaminantes, os quais são lançados diariamente em efluentes, solo ou pelo ar (COSTA et al., 2008).

Para realização do monitoramento toxicológico ambiental, muitas vezes, são utilizados ensaios com organismos vivos. O biomonitoramento se deu, nos EUA, a partir de 1984, quando houve uma implementação pela Agência Americana de proteção Ambiental de testes ecotoxicológicos de qualidade de água utilizando organismos vivos. Na mesma época houve uma consonância para utilização do biomonitoramento também nos países europeus (MAGALHÃES, FERRÃO FILHO, 2008).

A utilização de organismos vivos nos ensaios de ecotoxicidade pode ser conhecida como monitoramento toxicológico, bioensaio ou biomonitoramento (MAGALHÃES, FERRÃO FILHO, 2008). Apesar de serem capazes de sugerir possíveis respostas dos organismos humanos frente a determinados agentes tóxicos, os testes de biomonitoramento não fornecem resultados de completa extrapolação (COSTA et al., 2008). Ainda assim, os testes ecotoxicológicos com microorganismos vivos mostram características importantes de

toxicidade de agentes diversos, tais como limite de tolerância, alterações fisiológicas, morfológicas e comportamentais (MAGALHÃES, FERRÃO FILHO, 2008).

No Brasil, o biomonitoramento tem como marco uma iniciativa em 1975, na qual se utilizou peixes para padronização de testes de toxicidade. Estes estudos foram desenvolvidos pela agência de saneamento de água e esgoto paulista e por agências de padronização de normas técnicas (ZAGATTO E BERTOLETTI, 2008).

Ao longo dos anos, os testes de biomonitoramento vêm evoluindo para o uso de diferentes microorganismos. No Brasil, a Agência Brasileira de normas técnicas (ABNT) e a Companhia Ambiental do Estado de São Paulo (CETESB) são duas das primeiras agências que publicaram normativas, que envolveram o uso de algas, microcrustáceos e peixes, todos de águas doce ou salgada (ZAGATTO E BERTOLETTI, 2008).

Os resultados de ensaios químicos e físicos por si só não fornecem informações sobre o impacto ambiental que determinado agente pode causar. Neste sentido, o biomonitoramento, também conhecido como monitoramento biológico, pode ser considerado um meio ideal para gestão integrada de recursos hídricos, pois fornece resumo das condições gerais, que supera a deficiência de métodos meramente físicos ou químicos de avaliação toxicológica. Estes dois últimos passam a ser complementares ao biológico (BERE; TUNDISI, 2010).

A escolha do organismo dependerá de fatores como: disponibilidade, representatividade no ecossistema avaliado, conhecimento de seu comportamento e estrutura, custo, importância comercial e facilidade de cultivo. Não se tem conhecimento de espécies que reúnam todas as características desejadas para um objeto de estudo, sendo assim, deve-se escolher aquele que reunir o máximo de características e que represente melhor os efeitos causados por um agente tóxico em um determinado ecossistema (MAGALHÃES, FERRÃO FILHO, 2008).

No quadro 1, estão relacionadas alguns exemplos de padronização de testes de ecotoxicidade por agências reguladoras brasileiras, com seus respectivos microorganismos.

Quadro 1. Ensaio e normas brasileiras para padronização de testes ecotoxicológicos.

Ensaio	Legislação
Toxicidade Aguda com <i>Daphnia similis</i> Claus, 1876;	ABNT NBR 12713 ^{1,2} CETESB L 5.018 ^{1,2}
Parte I (Sistema estático) - Ensaio de Toxicidade Aguda com peixes	ABNT NBR 12714 ¹ CETESB L 5.019 ¹
Parte II (Sistema semi-estático) - Ensaio de Toxicidade Aguda com peixes	ABNT NBR 12715 ¹ CETESB L 5.019 ¹
Parte III (Sistema de fluxo contínuo) - Ensaio de Toxicidade Aguda com peixes	ABNT NBR 12716 ¹ CETESB L 5.019 ¹
Toxicidade com <i>Chlorellavulgaris</i>	ABNT NBR 12648 ^{1,2} CETESB L 5.020 ^{1,2}
Toxicidade crônica com <i>Ceriodaphnia dubia</i> Richard 1894	ABNT NBR 13373 ² CETESB L 5.022 ^{1,2}
Toxicidade aguda <i>Photobacteriumphosphoreum</i>	CETESB L 5.227 ¹
Toxicidade aguda <i>Spirillumvolutans</i>	CETESB L 5.228 ^{1,2}
Água do mar. Teste de toxicidade crônica <i>Lytechinusariegatus</i> Lamarck, 1816	CETESB L 5.250 ¹
Água do mar. Teste de toxicidade aguda com <i>Mysidopsisjuniae</i> Silva, 1979	CETESB L 5.251 ¹
Toxicidade aguda <i>Vibriofischeri</i>	CETESB L 5.227 ²
Toxicidade crônica <i>Scenedesmussubspicatus</i> , <i>Pseudokirchneriellasubcapitata</i>	ABNT NBR 12648 ² CETESB L 5.020 ²
Toxicidade aguda <i>Artemia salina</i>	CETESB L 5.021 ²
Toxicidade aguda <i>Daniorerioe Pimephalespromelas</i>	ABNT NBR 15088 ² CETESB L 5.019 ²

Fonte:¹ Magalhães e Ferrão Filho (2008); ² Costa et al. (2008).

Dentre os diferentes estudos em seres vivos animais que testam a toxicidade de diferentes moléculas, tem-se utilizado o crustáceo *Artemia salina*, um artrópode primitivo que se desenvolve em lagos salgados (SIQUEIRA et al., 1998; DUMITRASCU, 2011; ALAS et al., 2017).

3.2 *Artemia salina*

Trata-se de um microcrustáceo, da família *Artemiidae*, com origem de 100 milhões de anos. Apesar de ser endêmica no mediterrâneo, pode ser encontrada em todos os continentes. Sua alimentação baseia-se em algas, protozoários e detritos. Descrita como um crustáceo filtrador de ecossistemas talássicos, que se alimenta principalmente de bactérias, algas unicelulares, pequenos protozoários e detritos que são dissolvidos no meio (NETO, 2013).

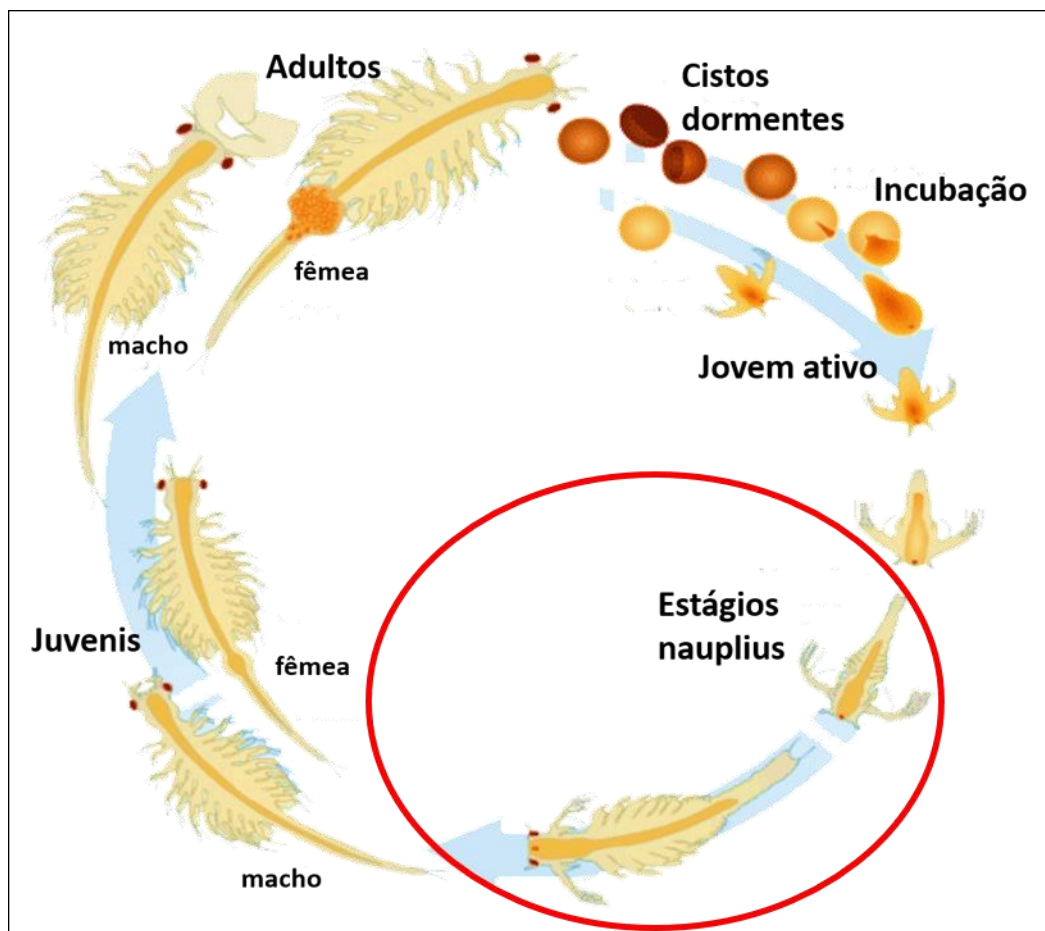
Este organismo apresenta um ciclo biológico complexo que começa com a incubação de cistos dormentes conforme ilustrado na figura 1. A hidratação deve causar incubação dentro de 48 horas, iniciando a primeira fase larval da *A. salina*. Os cistos absorvem a água de maneira rápida e fácil. Se as condições ambientais forem favoráveis, como a temperatura adequada, salinidade, luz solar e oxigênio os embriões internos começam o desenvolvimento.

O corpo é composto por três segmentos: cabeça, tórax e abdome. A espécie apresenta dimorfismo sexual, as principais diferenças morfológicas entre machos e fêmeas são a largura do terceiro segmento abdominal, comprimento total e o comprimento do abdome. O macho adulto atinge de 8 a 10 mm de comprimento, e a fêmea de 10 a 12 mm (DUMITRASCU, 2011).

As condições propícias de vida são calor, presença de luz solar e uma ampla faixa de concentração de sal. Apesar disso, são seres que sobrevivem em águas com baixas dosagens de oxigênio, podendo sobreviver a períodos de seca em estado de cisto, voltando a se reproduzir em situações propícias. Quando há melhora das condições ambientais, os cistos eclodem em larvas ou náuplias, as quais vão se transformar em adultos reprodutivos. Os cistos resistem até temperaturas de 80 °C. Porém, uma vez hidratados, morrem a temperaturas inferiores a 0 °C e superiores a 40 °C. A faixa de sobrevivência das náuplias gira em torno de 5 ppt e 70 ppt de salinidade (DUMITRASCU, 2011).

De acordo com das condições do meio, podem ser ovovivíparos (menos de 150 ppt de salinidade) ou ovíparos (150-200 ppt de salinidade). Serão ovíparos em condições com baixo oxigênio e alta salinidade, ou flutuações de oxigênio, ou presença de alimentos ricos em ferro. Contudo na presença de alto teor de oxigênio, baixa salinidade, ou em pequenas flutuações de oxigênio ou na presença de alimentos com baixo teor de ferro serão ovovivíparos (DUMITRASCU, 2011).

Figura 1. Esquemática do ciclo de vida da *Artemia salina*. Polígono em vermelho evidenciando o estágio de vida do organismo utilizada na pesquisa.



Fonte: Genetic Science Learning Center– Adaptado, 2017.

O náuplio recém-nascido trata-se, essencialmente, de uma cabeça nadadora com um tronco pequeno e relativamente pouco desenvolvido. Dentro de alguns dias, seu sistema digestivo amadurece e começa a se alimentar. Usa suas antenas para nadar e empurrar a comida até a boca, se alimentando de alga, bactéria e detrito que encontrar na água.

Quando o náuplio continua a se desenvolver, seu tronco cresce mais e começa a crescer uma série de membros semelhantes a remos chamados toracópodes. Os membros do

tronco (toracópodes) tornam-se totalmente funcionais e assumem os empregos de alimentação, natação e respiração. As antenas tornam-se menores e não ajudam mais o camarão de água salgada a nadar ou a se alimentar, como faziam durante os estágios nauplários. (Os estágios juvenis do desenvolvimento do camarão de água salgada também são chamados de fases pós-larvas). À medida que crescem, os machos e as fêmeas começam a parecer diferentes um do outro.

O náuplio sofre transformação cerca de 15 vezes antes de se tornar um adulto, entre duas e três semanas. A *A. salina* fêmeas são ligeiramente maiores que os machos e os seus sacos são facilmente visíveis a olho nú. Os machos não têm sacos de ninhada, e eles têm grandes grampos saindo do topo de suas cabeças. Durante o acasalamento, o macho agarra a fêmea com seus claspers. Ele fertiliza os óvulos da fêmea depositando o esperma nela. O adulto vive uma média de quatro a cinco semanas (CARPELAN, 1957).

O macho pode ter até 10mm, enquanto a fêmea pode atingir 12mm. Seu corpo é dividido em cabeça, tórax e abdômen, podendo apresentar 3 olhos e 11 pares de pernas na idade adulta, retratado na figura 2. Podem variar de verde a vermelho dependendo das concentrações salinas do meio. Há ao redor da boca um cérebro, assim como a maioria dos invertebrados (DUMITRASCU, 2011). No quadro 2 demonstra a classificação científica da *Artemia salina*.

Quadro 2. Classificação científica da *Artemia salina*.

Reino	Animália
Filo	Artrópode
Sub-filo	Crustáceo
Classe	Branchiopoda
Ordem	Anostraca
Família	Artemiidae
Gênero	Artemia
Espécie	<i>Artemia salina</i> (Linnaeus, 1758)

A importância da espécie reside no fato de ser utilizada na aquicultura como alimento para algumas espécies de peixes e crustáceos devido ao seu alto valor nutricional (DUMITRASCU, 2011; ALAS et al., 2017). Sua resiliência justifica a utilização nos estudos de toxicidade de vários produtos químicos (DUMITRASCU, 2011).

Figura 2. Representação fotográfica da *Artemia salina*.



Fonte: Dumitrascu (2011).

A *Artemia salina* é amplamente utilizada como um bioindicador de toxicidade, devido à alta sensibilidade no aumento da concentração de partículas, por ser um organismo filtrador, que apresenta alterações claras diante de baixas variações ambientais (NETO, 2013). Estudos tem mostrado que é um organismo com boa correlação com a atividade citotóxica contra tumores e capaz de detectar compostos bioativos em plantas (AMARANTE et al., 2011, SIQUEIRA et al., 1998). Outras atividades que têm sido testadas são, entre outras, antifúngica, viruscida, antimicrobiana, parasiticida, tripanossomicida. O uso deste microcrustáceo é considerado como uma boa alternativa pois estes bioensaio são de baixo custo, rápidos e não exigem técnicas assépticas (SIQUEIRA et al., 1998).

Os objetivos dos testes de toxicidade são avaliar ou prever efeitos tóxicos em sistemas biológicos e dimensionar a toxicidade relativa de determinada substância (BRASIL, 2002). A *Artemia salina*, devido às características já descritas, configura como um organismo promissor na obtenção de resultados rápidos e confiáveis.

Para Amarante et al. (2011), o “teste de toxicidade contra a *Artemia salina* é um ensaio biológico considerado como uma das ferramentas mais utilizadas para a avaliação preliminar de toxicidade”. Neste sentido, os autores testaram a toxicidade de uma planta muito utilizada na medicina tradicional amazônica, a *Montrichardia linifera*, para conhecer sua ação contra o parasita causador da malária. Os testes foram realizados utilizando extrato hexânico e o etanólico do caule da macrófita aquática em teste. O extrato etanólico apresentou alta toxicidade para *Artemia salina* e moderada atividade antiplasmódica. Diferentes concentrações foram testadas, sendo que em duas, foi observado alto potencial antimalárico.

Andrade et al. (2013) levantaram uma questão relativa à investigação da toxicidade de produtos em escala nano. Realizaram um estudo com objetivo analisar toxicidade de nano fibrilas de celulose, produzidas a partir do resíduo das bainhas da palmeira *Bactris gasipaes*. Para o bioensaio utilizaram a *Artemia salina*. As amostras de palmeira foram obtidas em uma agroindústria de palmito pupunha, do município de Morretes, no Paraná. Diferentes concentrações de nano fibrilas foram testadas nos tempos 24, 48 e 72 horas. Em pequenas concentrações, até 50 g/L, o material produzido não apresentou efeitos tóxicos sobre o microcrustáceo. Os autores consideraram que estas baixas concentrações serviram como nutrientes para a *Artemia salina*, já em altas concentrações interferiram no metabolismo e resultaram em morte.

Siqueira et al. (1998) utilizaram a *Artemia salina* como organismo teste para conhecer os efeitos de toxicidade de substâncias biologicamente ativas da *Unonopsis lindmanii* – *Annonaceae*. Os extratos foram obtidos por meio do processamento de folhas, cascas do caule e sementes da planta. A maior toxicidade ocorreu com amostras que continham um alcaloide chamado oxaporfínico liriodenina. Este achado corroborou com estudos revisados pelos autores, que mostraram que o alcalóide testado apresentou atividade citotóxica contra células tumorais de pulmão e de cólon de útero, evidenciando uma alta viabilidade e confiabilidade dos testes em *Artemia salina*.

Costa et al. (2008) apresentaram a *Artemia salina* como uma espécie bastante utilizada para testes de toxicidade devido sua facilidade de aquisição, baixo custo, fácil manipulação e rápido desenvolvimento até fase de testes. Segundo os autores, os organismos ficam aptos a testes em 1 ou 2 dias após os ovos secos terem sido expostos à água do mar ou ambiente simulado a uma temperatura de 23°. A facilidade de aquisição se deve ao fato de serem encontrados em abundância em lojas de produtos para aquários, pois são alimentos comercializados para peixes. Na legislação há uma padronização do uso de *Artemia salina* para testes de efeitos agudos regidas pela norma técnica CETESB, L5.021.

3.3 AVALIAÇÃO DE ECOTOXICIDADE

Experimentos laboratoriais podem simular situações em ambiente real, e demonstrar como ocorrem as interações entre os agentes tóxicos e microrganismos nativos, dos diferentes ambientes aquáticos. Em um momento de questionamentos e fortes críticas no uso de animais

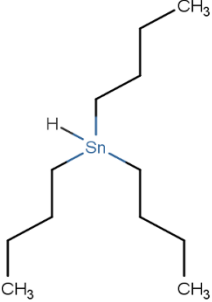
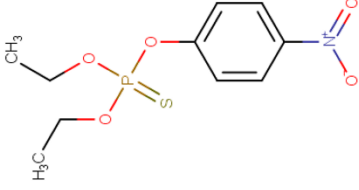
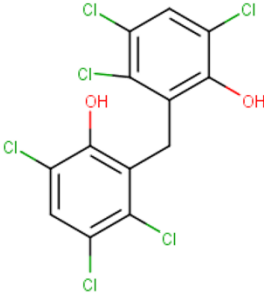
em experimentos, faz-se necessário a utilização de tecnologias alternativas na pesquisa de predição de efeitos biológicos diante de novos tóxicos (CZARIN et al., 2004).

O efeito de substâncias tóxicas sobre a biota aquática têm sido estudados, por meio de testes de toxicidade com organismos de águas continentais, estuarinas e marinhas, em condições laboratoriais e/ou de campo (BERTOLETTI, 1990). O objetivo é estabelecer limites aceitáveis para presença de substâncias químicas e avaliar o impacto de poluentes sobre os organismos dos corpos receptores (CONAMA, 2005). A Figura 3 demonstra exemplos de agrotóxicos já submetidos a testes laboratoriais.

A *A. salina* é amplamente utilizada como um bioindicador de toxicidade, devido à alta sensibilidade no aumento da concentração de partículas, por ser um organismo filtrador, que apresenta alterações claras diante de baixas variações ambientais (NETO, 2013). Os testes de toxicidade são executados com a finalidade de avaliar ou prever efeitos tóxicos em sistemas biológicos e dimensionar a toxicidade relativa das substâncias (BRASIL, 1989; IBAMA, 1996; BRASIL, 2002). Devido as características já descritas do microcrustáceo *Artemia salina*, tal organismo se torna promissor na obtenção de resultados rápidos e confiáveis.

Neste sentido, a tecnologia da informática pode se apresentar como uma ferramenta na redução do uso de animais em experimentos minimizando danos biológicos. Diante da perspectiva destes avanços, a inteligência artificial têm sido amplamente utilizada, para encontrar relações quantitativas entre estrutura química e toxicidade (DOBCHEV; PILLAI; KARELSON, 2014; MITCHELL, 2014). Sendo assim, os métodos computacionais têm sido reconhecidos como alternativas confiáveis, rápidas e de baixo custo para a avaliação da toxicidade de produtos químicos (CHERKASOV et al., 2014).

Figura 3. Exemplos de representação da forma estrutural de moléculas de agrotóxicos.

		
<p>Nome comercial: Hidreto de tributiltina</p> <p>Classificação toxicológica: Classe I. Muito tóxico para os organismos aquáticos com efeitos duradouros.</p> <p>Uso: Inseticida</p> <p>LC₅₀: 0,04141 ug/L</p> <p style="text-align: right;">A</p>	<p>Nome comercial: Parationa</p> <p>Classificação toxicológica: Casse I. Altamente Tóxico para organismos aquáticos, especificamente Crustáceos.</p> <p>Uso: Inseticida</p> <p>LC₅₀: 110 ug/L</p> <p style="text-align: right;">B</p>	<p>Nome comercial: Triclorofenol</p> <p>Classificação toxicológica: Classe I. Toxicidade para organismos aquáticos.</p> <p>Uso: Inseticida, fungicida, herbicida</p> <p>LC₅₀: 0,0081 ug/L</p> <p style="text-align: right;">C</p>

Diante da perspectiva dos avanços tecnológicos e as problemáticas ambientais, abordagens de aprendizado de máquina, o qual faz parte do campo da inteligência artificial, têm sido amplamente utilizadas para encontrar relações quantitativas entre estrutura química e toxicidade a partir de conjuntos de dados de compostos com toxicidade experimental determinada (DOBCHEV; PILLAI; KARELSON, 2014; MITCHELL, 2014).

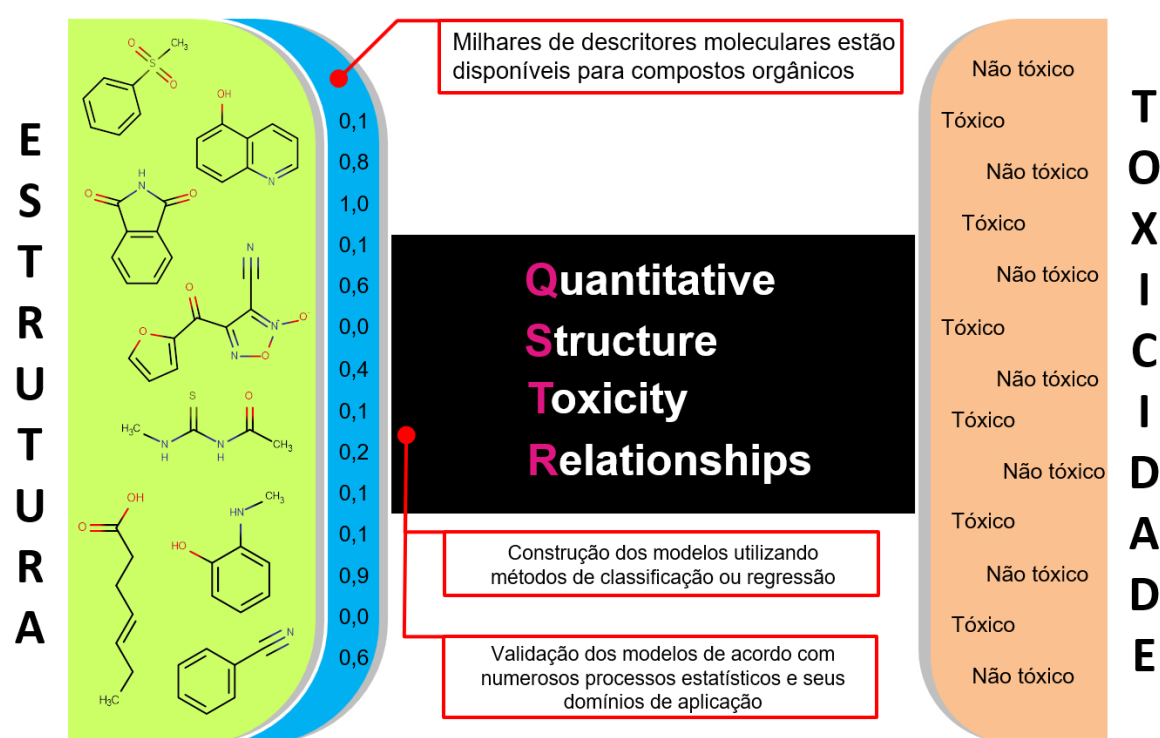
3.4 TOXICOLOGIA COMPUTACIONAL

Métodos computacionais têm sido reconhecidos como ferramentas alternativas confiáveis, rápidas e de baixo custo para a avaliação da toxicidade de produtos químicos. Nos últimos dez anos, diversas abordagens como tais como o aprendizado de máquina, que é um dos componentes mais importantes da inteligência artificial, têm sido amplamente utilizadas para encontrar relações quantitativas ou qualitativas entre estrutura química e a toxicidade a partir de conjuntos de dados de compostos com toxicidade experimental pré-determinada disponíveis na literatura ou em bancos de dados particulares (CHERKASOV et al., 2014).

Metodologicamente, a modelagem dos parâmetros toxicológicos pode ser representada como um processo de três partes como representado na Figura 4. Inicialmente, as estruturas químicas de moléculas (*e.g.*, agrotóxicos) são codificadas em descritores moleculares (variáveis independentes). Em seguida, métodos de aprendizado de máquina são utilizados

para estabelecer relações quantitativas entre os descritores moleculares e propriedade toxicológica estudada (variável dependente). Esta etapa envolve a descoberta empírica de uma função que irá estabelecer pesos aos descritores moleculares, ajustando-os a equação $P_t = k' (D_1, D_2, \dots, D_n)$, na qual P_t é a propriedade toxicológica das moléculas, D_1, D_2, \dots, D_n são os descritores moleculares, e k' é um peso estabelecido empiricamente pelo algoritmo selecionado (CHERKASOV et al., 2014; TROPSHA, 2010). Ao final deste processo, a preditividade do modelo obtido é calculada utilizando métricas apropriadas, as quais serão avaliadas a habilidade do modelo em prever corretamente a toxicidade de compostos avaliados experimentalmente (CHERKASOV et al., 2014; TROPSHA, 2010). Uma vez validado, o modelo gerado representa uma ferramenta valiosa para a priorização de novos produtos químicos em estágios de desenvolvimento e avaliação preditiva de toxicidade para produtos aprovados (CHERKASOV et al., 2014).

Figura 2. Principais etapas envolvidas na construção e validação de modelos de QSTR



Fonte: (Adaptado de TROPSHA, 2010).

3.5 DESCRITORES MOLECULARES

O descritor molecular é o resultado final de um procedimento lógico e matemático que codifica a estrutura química de um composto em um número útil (BUCKLE et al., 2013;

TODESCHINI; CONSONNI, 2000, 2009). Atualmente, milhares de descritores moleculares estão disponíveis (LANDRUM, 2014; TODESCHINI; CONSONNI, 2000, 2009; YAP, 2011) para a construção de modelos de QSTR.

Estes podem ser classificados em unidimensionais (1D), baseados em propriedades físico-químicas e na fórmula molecular (*e.g.* refratividade molar, logP e massa molecular); bidimensionais (2D), que descrevem propriedades que podem ser representadas como ligações e átomos (*e.g.* número de átomos de nitrogênio e oxigênio, número de ligações e índices de conectividade); e tridimensionais (3D), que dependem da conformação das moléculas (*e.g.* área de superfície acessível ao solvente e volume molecular) (XUE; BAJORATH, 2000).

Os descritores moleculares também podem ser classificados de acordo com o tipo de informação química: (i) impressões digitais moleculares (do inglês molecular *fingerprints*), que calculam a presença ou frequência de determinados fragmentos na molécula (*e.g.* *fingerprints* Morgan e FeatMorgan); (ii) quanto-mecânicos (*e.g.* energia dos orbitais moleculares); (iii) eletrostáticos (*e.g.* índices de polaridade e carga parciais); (iv) constitucionais (*e.g.* peso molecular, números de átomos e ligações); (v) topológicos (*e.g.* índices de Randić); e (iv) geométricos (*e.g.* volume molecular, área de superfície polar) (KOCKAR et al., 2010; ROGERS; HAHN, 2010; TODESCHINI; CONSONNI, 2009).

3.6 APRENDIZADO DE MÁQUINA

O aprendizado de máquina é um ramo da inteligência artificial que estuda o desenvolvimento de sistemas capazes de aprender de forma automática. O sistema de aprendizado representa um algoritmo que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores. Esse aprendizado é construído em três estágios: (i) representação dos dados; (ii) otimização da hipótese; e (iii) generalização. Por exemplo, durante a construção de modelos de QSTR, os compostos podem ser representados como descritores moleculares. Em seguida, uma hipótese é gerada na tentativa de estabelecer uma relação dos descritores com a atividade biológica e essa relação é otimizada.

3.7 BOAS PRÁTICAS DE DESENVOLVIMENTO E VALIDAÇÃO EM QSTR

Por se tratar de um campo de pesquisa relativamente novo, nos últimos 15 anos várias diretrizes e recomendações foram desenvolvidas para auxiliar a construção e validação de modelos de QSTR (CHERKASOV et al., 2014; DEARDEN; CRONIN; KAISER, 2009; TROPSHA, 2010). A principal delas, publicada no ano de 2004 pela OECD sugere que a modelagem de QSTR siga cinco princípios: (i) ter propriedade biológica definida; (ii) utilizar algoritmos não ambíguos; (iii) ter domínio de aplicabilidade (DA) definido; (iv) apresentar robustez e preditividade apropriadas; e (v) realizar a interpretação mecanística dos modelos, ou seja, encontrar relações entre os descritores e a atividade biológica, visando a melhor compreensão do mecanismo de ação de uma estrutura química ou aprofundar o conhecimento biológico sobre a propriedade em estudo (OECD, 2004).

4. MATERIAIS E MÉTODOS

Todas as etapas *in silico* desenvolvidas neste estudo foram executadas em Python v.3.6 (<https://www.python.org>). O *frame* utilizado é completo, como o módulo para preparação de dados, balanceamento de conjuntos de dados, construção e validação de modelos QSTR e interpretação mecanística.

4.1 INTEGRAÇÃO, PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS

Todos os compostos com dados de Dose Letal DL_{50} e Concentração Letal - CL_{50} para *A. salina* testados na *in vivo* em *Artemia salina* foram extraídos dos bancos de dados ECOTOX (<https://cfpub.epa.gov/ecotox/>) e ChEMBL (<https://www.ebi.ac.uk/chembl/>). Em seguida, compostos com dados de toxicidade expressos na escala de peso molecular (mg/mL, µg/mL, etc.) foram convertidos para escala molar (µM/mL). Ao final deste processo, compostos com $LC_{50} \leq 10$ µM/mL foram classificados como tóxicos, ao passo que compostos com $LC_{50} > 10$ µM/mL foram classificados como não tóxicos, seguindo critérios de determinação proposto pela *Chemical Hazard Classification and Labeling: comparison of opp requirements and theghs*.

Após a preparação dos dados, todos os compostos foram cuidadosamente padronizados de acordo com o protocolo estabelecido por Fourches e colaboradores (ALVES et al., 2015b; FOURCHES; MURATOV; TROPSHA, 2015, 2016). Brevemente, hidrogênios explícitos foram adicionados enquanto que polímeros, sais, metais, compostos organometálicos e misturas foram removidos. Em paralelo, quimiotipos específicos como anéis aromáticos e grupos nitro foram normalizados. Todas estas etapas foram realizadas utilizando o programa Standardizer (v.16.9.5.0, ChemAxon, Budapest, Hungary, <http://www.chemaxon.com>). Em seguida, todos os conjuntos de dados foram submetidos a um processo de análise e remoção de duplicatas seguindo os seguintes critérios: (i) duplicatas com propriedades toxicológicas discordantes foram excluídas; e (ii) duplicatas com propriedades toxicológicas concordantes tiveram uma das entradas mantida no conjunto de dados e as demais excluídas.

4.2 BALANCEAMENTO DO CONJUNTO DE DADOS

O conjunto de dados foi balanceado utilizando uma estratégia de sub-amostragem desenvolvida pelo nosso grupo (NEVES et al., 2016). Inicialmente, foi calculada a distância Euclidiana da matriz contendo os compostos da classe minoritária (tóxicos), representados por impressões digitais moleculares FeatMorgan (seção 4.3.), em relação a cada composto da classe majoritária (não tóxicos), utilizando o método de *k*NN (do inglês *k-nearestneighbors*). Em seguida, amostras da classe majoritária foram linearmente extraídas ao longo de toda faixa de *k*-distâncias (compostos representativos de todas as faixas de distância), compondo um conjunto final balanceado em relação à classe minoritária.

4.3 DESCRITORES MOLECULARES

As impressões digitais moleculares circulares do tipo Morgan (ECFP) e FeatMorgan (FCFP) foram calculadas no programa de código aberto RDKit (<http://www.rdkit.org>, (RINIKER; LANDRUM, 2013b)). Ambas as impressões digitais foram geradas com tamanho de raio variando entre 2–6 e com comprimento de até 2.048 bits.

4.4 CONSTRUÇÃO DOS MODELOS

Os modelos de QSTR foram construídos combinando impressões digitais moleculares com o algoritmo de classificação *Random Forest* e método *Support Vector Machine* (SVM) pacote de código aberto Scikit-learn (<http://scikit-learn.org>) (PEDREGOSA et al., 2012). Os parâmetros de busca da grade foram definidos utilizando combinações de árvores (100, 150, 200, 250, 300, 350, 400, 450, e 500) e o coeficiente kappa de Cohen's como função de pontuação do estimador. Os parâmetros de busca da grade do SVM utilizados foram: C (1, 3, 5, 7, 10, 15, 20, 50, 100, 1000), gama (0,01, 0,001, 0,0001, 0,00001), kernel (função de base radial, sigmoide ou linear) e o coeficiente kappa de Cohen's como função de pontuação do estimador.

4.5 VALIDAÇÃO CRUZADA EXTERNA DE 5-FOLDS (5FECV)

Os conjuntos de dados foram divididos aleatoriamente em cinco subconjuntos de igual tamanho; então um desses subconjuntos (20% de todos os compostos) foi considerado como conjunto teste e as demais partes juntas formaram o conjunto de treinamento (80% do conjunto completo). Este procedimento foi repetido cinco vezes, permitindo que cada um dos cinco subconjuntos fosse usado como conjunto de validação e teste. Os modelos foram construídos utilizando apenas usando o conjunto de treinamento. É importante ressaltar que os compostos no conjunto teste momentâneo não foram utilizados para construir os modelos (ALVES et al., 2015a, 2015b).

4.6 AVALIAÇÃO DE PREDITIVIDADE DOS MODELOS

O desempenho preditivo dos modelos binários gerados foi avaliado utilizando as métricas calculadas, conforme as equações para sensibilidade (SE), especificidade (SP), taxa de classificação correta (CCR), valor preditivo positivo (PPV) e valor preditivo negativo (NPV), conforme descrito abaixo.

$$SE = \frac{VP}{VP + FN} \quad (1)$$

$$SP = \frac{VN}{VN + FP} \quad (2)$$

$$CCR = \frac{SE + SP}{2} \quad (3)$$

$$PPV = \frac{VP}{VP + FP} \quad (4)$$

$$NPV = \frac{VN}{VN + FN} \quad (5)$$

Aqui, VP representa o número de verdadeiros positivos, VN o número de verdadeiros negativos verdadeiros, FP o número de falsos positivos, FN é o número de falsos negativos e N é o número total de compostos. A SE representa uma medida de correção na predição de produtos químicos tóxicos. A SP caracteriza uma medida de correção na predição de produtos químicos não tóxicos. Já o CCR é a média aritmética dos valores de SE e SP. Por fim, PPV e NPV foram calculados para estimar a probabilidade de acerto durante a predição um novo

composto tóxico ou não tóxico. Valores de CCR, SE, SP, PPV e NPV maiores que 0,65 indicam que os modelos são preditivos.

Além das métricas acima, o κ será utilizado para comparar o quão bem as predições concordam com os valores experimentais.(COHEN, 1960; VIERA; GARRETT, 2005). Ele será calculado utilizando as equações a seguir.

$$\text{Pr}(a) = \frac{\text{VP} + \text{VN}}{\text{N}} \quad (6)$$

$$\text{Pr}(e) = \frac{(\text{VP} + \text{FP}) \times (\text{VP} + \text{FN}) + (\text{VN} + \text{FN}) \times (\text{VN} + \text{FP})}{\text{N}} \quad (7)$$

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (8)$$

Nas equações acima, $\text{Pr}(a)$ representa a concordância relativa observada entre a predição realizada pelo modelo e os valores conhecidos e $\text{Pr}(e)$ representa a probabilidade hipotética de concordância. Valores de κ maiores que 0.6 indicam boa concordância entre os valores preditos e experimentais.

4.7 DOMÍNIO DE APLICABILIDADE

O DA foi estimado com base na distância Euclidiana entre os compostos do conjunto de modelagem (5FECV) e cada compostos no subconjunto de dados de validação externa. Na presente proposta, o DA será definido como um limiar de distância L_D entre um composto submetido a uma predição e seu vizinho mais próximo no conjunto modelagem. O DA será calculado utilizando a equação (9):

$$L_D = \bar{y} + Z\sigma \quad (9)$$

Na qual \bar{y} é a distância Euclidiana média de k vizinhos mais próximos dentro do conjunto de modelagem, σ é o desvio padrão dessas distâncias Euclidianas e Z (0,5) um parâmetro arbitrário para controlar o nível de significância. Se a distância de um composto exceder o limiar estabelecido, a predição será considerada como menos confiável (TROP SHA, 2010).

4.8 INTERPRETAÇÃO MECANÍSTICA DOS MODELOS

Mapas de probabilidade foram gerados a partir do modelo mais preditivo para visualizar os átomos e fragmentos moleculares que contribuem para a toxicidade em *A. salina*. Nesta abordagem, o "peso" de um átomo foi considerado como a diferença de probabilidade obtida quando os *bits* da impressão digital molecular são removidos (RINIKER; LANDRUM, 2013a). Os pesos normalizados foram então usados para colorir os átomos em um mapa topográfico, no qual a cor roxa indica uma contribuição positiva e a cor verde indica uma contribuição negativa para a toxicidade.

5. RESULTADOS E DISCUSSÃO

5.1. PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS

Um tipo de prática bastante importante no campo da quimioinformática diz respeito ao preparo e padronização dos conjuntos de dados utilizados na modelagem de QSTR. Esta prática baseia-se no conceito “confiável, porém verificável” (do inglês, *trust but verify*). Geralmente, conjuntos de dados públicos contêm uma fração de moléculas compiladas erroneamente. Em termos gerais, esses dados "ruins" englobam entradas que contêm anotações de estereoisômeros pouco claras, estruturas químicas com problemas de valência, valores de toxicidade reportados erroneamente e registros duplicados (FOURCHES et al., 2010). No presente estudo, 725 moléculas com dados de LC₅₀ para *A. salina* foram preparadas, padronizadas e visualmente inspecionadas. Tal prática permitiu a identificação de 101 duplicatas e 31 compostos com dados toxicológicos inconclusivos.

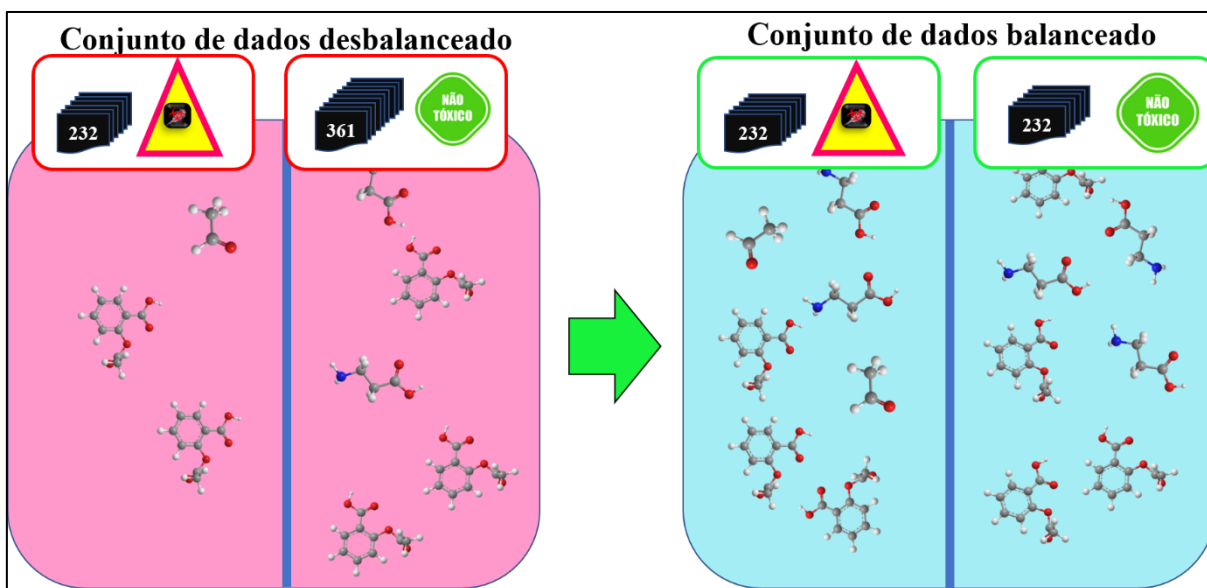
Os dados compilados a partir de diferentes condições experimentais (variância interlaboratorial e intralaboratorial, frequentemente associada aos resultados da ChEMBL e ECOTOX) geralmente apresentam certa discordância (FOURCHES; MURATOV; TROPSHA, 2015). O uso inadequado destes conjuntos de dados pode acrescentar ruído aos modelos de QSTR, portanto recomenda-se que todas as duplicatas com dados toxicológicos discordantes removidos (FOURCHES et al., 2010; FOURCHES; MURATOV; TROPSHA, 2015, 2016). Durante a análise de concordância, constatou-se que 100% das duplicatas tinham dados toxicológicos concordantes, ou seja, todas as duplicatas pertenciam a mesma classe (tóxico ou não tóxico), indicando a qualidade do conjunto de dados.

5.1.1 PREPARO E PADRONIZAÇÃO DO CONJUNTO DE DADOS

O conjunto de dados preparado apresentou classes com diferentes tamanhos (232 compostos tóxicos e 361 não tóxicos). Portanto, o conjunto de dados foi balanceado utilizando uma estratégia de sub-amostragem baseada na distância euclidiana de cada composto do conjunto teste (5FECV) em relação a todos os compostos no conjunto de treinamento (NEVES et al., 2016), uma vez que os métodos de aprendizado de máquina tendem a classificar amostrar como pertencentes a classe majoritária. Ao final do processo de balanceado, obteve-se um conjunto de dados com classes contendo o mesmo peso, 232

compostos tóxicos e 232 compostos não tóxicos, como ilustrado na Figura 5. Ao contrário dos métodos tradicionais de sub-amostragem que balanceiam aleatoriamente o conjunto de dados, esta estratégia manteve os compostos mais representativos no conjunto de dados balanceado, garantindo o máximo possível cobertura do espaço químico original.

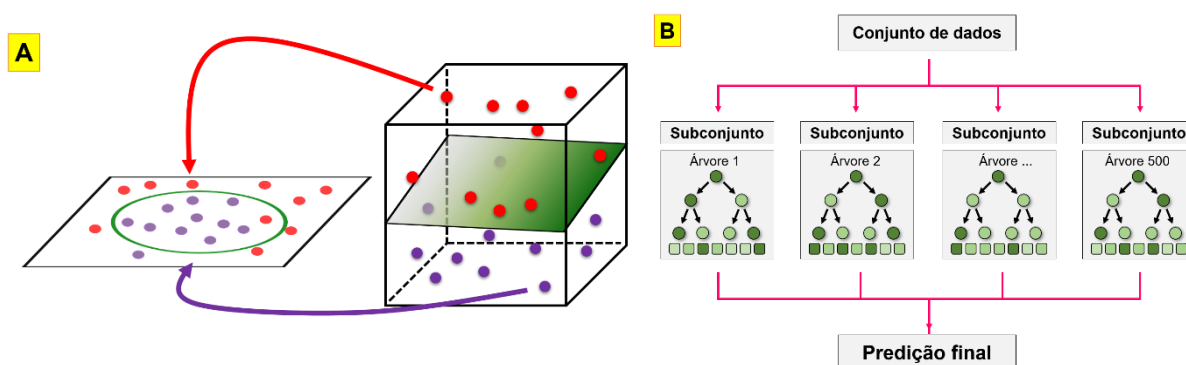
Figura 5: Figura esquemática do preparo e padronização do conjunto de dados.



5.1.2 MODELOS DE QSTR

No presente estudo, modelos binários de QSTR foram desenvolvidos para avaliação preditiva de ecotoxicidade em *A. salina* utilizando dois métodos de aprendizado de máquina: Random Forest e SVM, como representado na Figura 6.

Figura 6. Representação do plano de separação gerado por SVM (A), e do consenso das predições obtidas pelas árvores de decisão no método de Random Forest (B).



Fonte: (Adaptado de NEVES, 2015).

O método SVM, representado na figura acima, mapeou os dados em um espaço altamente dimensional, utilizando uma função de Kernel radial. Em seguida, o algoritmo encontrou um plano de separação ótima entre duas classes através da maximização da margem entre os pontos mais próximos, conhecidos como suporte de vetores.

O método de SVM utilizado mapeia os dados em um espaço altamente dimensional, utilizando uma função de Kernel radial. Em seguida, o algoritmo procura encontrar um plano de separação ótima entre duas classes através da maximização da margem entre os pontos mais próximos, conhecidos como suporte de vetores. Em outras palavras, um modelo obtido por SVM representa os compostos (descritores) como pontos num espaço e os mapeia de maneira que cada classe (i.e. tóxicos e não tóxicos) seja dividida por um espaço claro que seja tão amplo quanto possível (VAPNIK, 2000). Já o método Random Forest é uma técnica de classificação baseada na construção de diversas árvores de decisão. Os resultados de predição de cada árvore de decisão são agregados em uma abordagem de consenso para se obter uma predição final (BREIMAN, 2001).

A combinação das impressões digitais Morgan e FeatMorgan (raio 2: FeatMorgan_2, Morgan_2; raio 4: FeatMorgan_4, Morgan_4, raio 6: FeatMorgan_6, Morgan_6) com o método *Random Forest* levou a obtenção de modelos altamente preditivos, com valores de taxa de classificação correta (CCR) variando entre 0,81–0,82, sensibilidade (SE) de 0,78–0,81, especificidade (SP) de 0,82–0,86, valor preditivo positivo (PPV) de 0,82–0,85, valor preditivo negativo (NPV) de 0,79–0,82, coeficiente kappa de Cohen's (κ) de 0,63–0,65 e uma cobertura de 0,57–0,66 (Quadro 3). O aumento do raio (i.e. parâmetro que define tamanho dos fragmentos moleculares computado durante o cálculo de descritores) apresentou influência negativa sobre a taxa de acerto na predição dos compostos tóxicos (redução de 2% nas SEs, em relação ao melhor modelo) e influência positiva sobre a taxa de classificação de compostos não tóxicos (~2% nas SPs). A combinação das impressões digitais Morgan e FeatMorgan (raio 2: FeatMorgan_2, Morgan_2; raio 4: FeatMorgan_4, Morgan_4, raio 6: FeatMorgan_6, Morgan_6) com o método SVM também levou a obtenção de modelos de QSTR preditivos, com valores de (CCR) variando entre 0,79–0,81, (SE) de 0,81–0,85, (SP) de 0,76–0,80, (PPV) de 0,77–0,80, valor (NPV) de 0,80–0,84, (κ) de 0,59–0,62 e uma cobertura de 0,57–0,66, como expresso no quadro 3.

Entretanto, o modelo de QSTR mais preditivo deste estudo foi obtido através da melhor modelo foi construído com combinação das impressões digitais impressão digital FeatMorgan_2 e Random Forest, uma vez que apresentou o maior CCR (0.82) e κ (0.65).

Morgan é uma impressão digital circular construída pela aplicação do algoritmo Morgan a um conjunto de estruturas químicas 2D fornecidas pelo usuário. O processo de geração de impressões digitais registra sistematicamente a vizinhança de cada átomo não-hidrogênio em múltiplas camadas circulares até um raio estabelecido. O Morgan captura informações atômicas altamente específicas, permitindo a representação de um grande conjunto de características estruturais precisamente definidas.

Tabela 3. Resultados estatísticos dos modelos de QSTR construídos para predição de ecotoxicidade em *A. salina*.

Modelo	CCR	SE	SP	PPV	NPV	κ	Cobertura
Floresta Randômica							
Morgan_2	0,82	0,81	0,83	0,82	0,81	0,64	0,61
Morgan_4	0,81	0,81	0,82	0,82	0,82	0,63	0,58
Morgan_6	0,81	0,79	0,84	0,83	0,80	0,63	0,57
FeatMorgan_2*	0,82	0,80	0,85	0,84	0,81	0,65	0,66
FeatMorgan_4	0,82	0,78	0,86	0,85	0,79	0,63	0,58
FeatMorgan_6	0,82	0,78	0,86	0,85	0,79	0,63	0,57
Máquina de Suporte de Vetores							
Morgan_2	0,79	0,81	0,77	0,78	0,81	0,59	0,61
Morgan_4	0,79	0,83	0,76	0,77	0,81	0,59	0,58
Morgan_6	0,81	0,81	0,80	0,80	0,80	0,60	0,57
FeatMorgan_2*	0,81	0,85	0,77	0,78	0,84	0,62	0,66
FeatMorgan_4	0,80	0,82	0,79	0,80	0,81	0,61	0,58
FeatMorgan_6	0,81	0,82	0,80	0,80	0,81	0,62	0,57

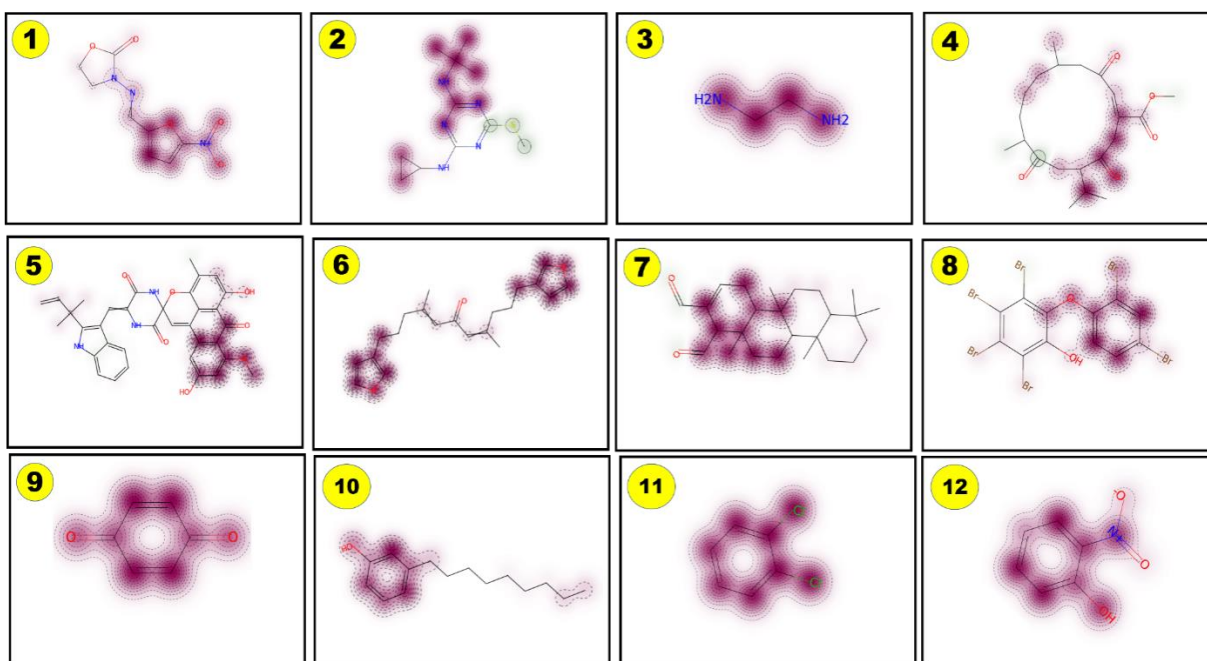
*Modelo de QSTR mais preditivo; CCR: taxa de classificação correta; SE: sensibilidade; SP: especificidade; PPV: valor preditivo positivo; NPV: valor preditivo negativo; κ : coeficiente kappa de Cohen's; Cobertura: porcentagem de compostos do conjunto teste dentro do DA.

5.1.3 ALERTAS ESTRUTURAIS

Após o processo de construção e validação, o modelo mais preditivo foi mecanisticamente interpretado utilizando os mapas de probabilidade predita.

Através desta abordagem foi possível identificar moléculas e fragmentos com contribuição positiva para a toxicidade em *A. salina* (alertas estruturais), tais como: 5-nitrofurano (1), etilenodiamina (3), quinona (9) e 1,2-diclorobenzeno (11). A lista de alertas estruturais está apresentada na Figura 7. Considerando o potencial destas moléculas e fragmentos para a toxicidade, é extremamente recomendável que essas subestruturas sejam evitadas durante o processo de desenvolvimento de novos agrotóxicos ou tenham aprovação de uso negada, caso estejam em fase de registro nos órgãos reguladores ambientais.

Figura 7. Alertas estruturais identificados a partir da interpretação mecanística do modelo FeatMorgan_2. Os fragmentos ou átomos coloridos em roxo possuem contribuição positiva para a toxicidade em *A. salina*.



6. CONCLUSÕES E PERSPECTIVAS

- Foi compilado, integrado e preparado um conjunto de dados de compostos de LC₅₀ para *A. salina* disponível na literatura;
- Todos os modelos de QSTR desenvolvidos foram robustos e preditivos;
- O aumento do raio das impressões digitais moleculares apresentou influência negativa sobre SE e influência positiva sobre SP dos modelos gerados utilizando o método Random Forest;
- Os modelos desenvolvidos com o método Random Forest apresentaram maior preditividade em comparação com o método SVM;
- O modelo de QSTR mais preditivo foi desenvolvido através da combinação de FeatMorgan_2 e Random Forest (CCR = 0,82; SE = 0,80; SP = 0,85; PPV= 0,84; NPV= 0,81; k= 0,65; Cobertura= 0,66);
- A interpretação mecanística do melhor modelo gerado (combinação de descritores Morgan_2 com método de aprendizado de máquina Random Forest) através dos mapas de probabilidade predita foi útil para compreender a relação entre grupos e/ou fragmentos estruturais com contribuição para toxicidade em *A. salina*;
- Os modelos gerados serão implementados em servidores online para que sejam utilizados por profissionais na avaliação preditiva de ecotoxicidade de substâncias ainda não testadas em *A. salina*;
- Os modelos gerados nesse trabalho foram utilizados na análise preditiva de toxicidade de dois compostos desenvolvidos com potencial agrotóxico. Esses dados foram finalizados e serão publicados em uma revista de alto impacto;
- Com base na metodologia compreendida e aplicada na elaboração dessa dissertação colaborei na escrita de um artigo, intitulado “Desenvolvimento de Modelos Computacionais para Avaliação Preditiva de Ecotoxicidade em Abelhas: Desafios Atuais”, já submetido e aprovado para publicação na Revista Fronteiras (B1).

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALAS, A.; KAYA, M.; OKTENER, A. Distribution and abundance of *Artemia salina* in the salt lake basin (Central Anatolia, Turkey). *Transylv. Rev. Syst. Ecol. Res.*; v. 19.2, 2017. <https://doi.org/10.1515/trser-2017-0011>
- ALVES, V. M. et al. Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization. *Toxicol Appl Pharmacol*, v. 284, n. 2, p. 273–280, 2015a. <https://doi.org/10.1016/j.taap.2014.12.013>
- ALVES, V. M. et al. Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. *Toxicology and Applied Pharmacology*, v. 284, n. 2, p. 262–272, abr. 2015b. <https://doi.org/10.1016/j.taap.2014.12.014>
- AMARANTE, C. B.; MÜLLER, A. H.; PÓVOA, M. M.; DOLABELA, M. F. Estudo fitoquímico biomonitorado pelos ensaios de toxicidade frente à *Artemia salina* e de atividade antiplasmódica do caule de aninga (*Montrichardia linifera*). *Acta Amazonica*. v. 41, n. 3, p. 431-4, 2011. <http://dx.doi.org/10.1590/S0044-59672011000300015>
- ANDRADE, D. R. M. et al. Avaliação toxicológica em *Artemia salina* de suspensão de nanofibrilas de celulose a partir do resíduo da pupunha. In: VII Workshop de Nanotecnologia Aplicada ao Agronegócio. Embrapa Instrumentação, São Carlos, 10 a 13 de junho de 2013.
- BERE, T.; TUNDISI, J. G. Biological monitoring of lotic ecosystems: the role of diatoms. *Braz. J. Biol.*, v. 70, n. 3, p. 493-502, 2010. <http://dx.doi.org/10.1590/S1519-69842010005000009>
- BERTOLETTI, E. Toxicidade e concentração de agentes tóxicos em efluentes industriais. *Revista Ciência e Cultura*, v. 43, n. 3/4, p. 271-277, 1990.
- BRASIL. Lei no 7.802, de 11 de julho de 1939. Presidência da Republica. Disponível em: < http://www.planalto.gov.br/ccivil_03/Leis/L7802.htm>. Acesso em: 02 jun. 2018.
- BRASIL. Decreto no 4.074, de 04 de janeiro de 2002. Presidência da Republica. Disponível em: < http://www.planalto.gov.br/ccivil_03/decreto/2002/d4074.htm>. Acesso em 02 jun. 2018.
- Breiman, L. *Machine Learning* (2001) 45: 5. <https://doi.org/10.1023/A:1010933404324>
- Buckle, D. R., Erhardt, P. W., Ganellin, C. R., Kobayashi, T., Perun, T. J., Proudfoot, J. and Senn-Bilfinger, J. (2013). Glossary of Terms Used in Medicinal Chemistry Part II. *Pure Appl Chem*. 48, 387–41. <https://doi.org/10.1351/PAC-REC-12-11-23>
- CARPELAN, L. H. Hydrobiology of the Alviso Salt Ponds. *Ecology*, v. 38, n. 3, p. 375, jul. 1957. <https://doi.org/10.2307/1929880>
- 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, 26 jun. 2014. <https://doi.org/10.1021/jm4004285>

COHEN, J. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, v. 20, p. 37–46, 1960. <https://doi.org/10.1177/001316446002000104>

COSTA, C. R. et al. A toxicidade em ambientes aquáticos: Discussão e métodos de avaliação. *Química Nova*, v. 31, n. 7, p. 1820–1830, 2008. <http://dx.doi.org/10.1590/S0100-40422008000700038>

CONAMA. Resolução n 357, 18 de março de 2005. *Diário Oficial*, n. 053, p. 58–63, 2005. Disponível em: < <http://www.mma.gov.br/port/conama/legiabre.cfm?codlegi=459>>. Acesso em: 04 jun. 2018.

CZARIN, K. C. C.; CORRÊA, C. L.; ZAMBRONE, F. A. D. Redução, refinamento e substituição do uso de animais em estudos toxicológicos: uma abordagem atual. *Braz Journal of Pharmaceut Sciences*, v. 40, n. 3, jul/set 2004. <http://dx.doi.org/10.1590/S1516-93322004000300004>

DEARDEN, J. C.; CRONIN, M. T. D.; KAISER, K. L. E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, v. 20, n. 3–4, p. 241–266, 2009. <https://doi.org/10.1080/10629360902949567>

DÍAZ-CRUZ, M. S.; BARCELÓ, D. Trace organic chemicals contamination in ground water recharge. *Chemosphere*, v. 72, n. 3, p. 333–42, 2008. <https://doi.org/10.1016/j.chemosphere.2008.02.031>

DOBCHEV, D.; PILLAI, G.; KARELSON, M. In silico machine learning methods in drug development. *Current Topics in Medicinal Chemistry*, v. 14, n. 16, p. 1913–1922, out. 2014. <https://doi.org/10.2174/1568026614666140929124203>

DUMITRASCU, M. Artemia salina. *Balneo Research Journal*, v. 2, n. 4, p. 119–122, 2011. <https://doi.org/10.12680/balneo.2011.1022>

FOURCHES, D. et al. 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. 1189–204, jul. 2010. <https://doi.org/10.1021/ci100176x>

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Curation of chemogenomics data. *Nature Chemical Biology*, v. 11, n. 8, p. 535–535, 1 ago. 2015. <https://doi.org/10.1038/nchembio.1881>

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. 1243–52, 25 jul. 2016. <https://doi.org/10.1021/acs.jcim.6b00129>

KOCKAR, H. et al. Co-Fe films: effect of Fe content on their properties. *Journal of nanoscience and nanotechnology*, v. 10, n. 11, p. 7639–42, nov. 2010. <https://doi.org/10.1166/jnn.2010.2761>

LANDRUM, G. RDKit: Open-Source Cheminformatics Software, 2014. Disponível em: < https://ptabdata.blob.core.windows.net/files/2016/PGR2016-00042/v137_1140%20-%20COR_RDKit%20Website.pdf>. Acesso em: 05 jun. 2018.

IBAMA. Portaria Normativa IBAMA nº 84, de 15 de outubro de 1996. Disponível em: <https://servicos.ibama.gov.br/ctf/manual/html/Portaria_84.pdf>. Acesso em: 02 jun. 2018.

MAGALHÃES, D. P.; FERRÃO FILHO, A. S. A ecotoxicologia como ferramenta no biomonitoramento de ecossistemas aquáticos. *Oecol. Bras.* v. 12, n. 3, p. 355-81, 2008.

MILLMAN, K. J.; AIVAZIS, M. Python for Scientists and Engineers. *Computing in Science & Engineering*, v. 13, n. 2, p. 9–12, mar. 2011. <https://doi.org/10.1109/MCSE.2011.36>

MITCHELL, J. B. O. Machine learning methods in chemoinformatics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 4, n. 5, p. 468–481, 1 set. 2014. <https://doi.org/10.1002/wcms.1183>

MONTANHA, F. P. et al. Degradação de ambientes aquáticos por exposição a compostos químicos. *Revista eletrônica de medicina veterinária*, v. 9, n. 17, p. 1–12, 2011. Disponível em: <http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/EYehChKPyqmghtA_2013-6-27-15-27-44.pdf>. Acesso em: 01 jun. 2018.

NETO, V. O. B. Análise in vitro do grau de toxicidade do extrato etanólico de *Azadiracta indica* A. juss em *Artemia salina* e em larvas infectantes de *Haemonchus contortus*, 2013. Disponível em: <http://www.cstr.ufcg.edu.br/grad_med_vet/mono_2013_2/monografia_vicente_oliveira_brito_netto.pdf>. Acesso em: 01 jun. 2018

NEVES, B. J. et al. Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening. *Journal of Medicinal Chemistry*, v. 59, n. 15, p. 7075–7088, 11 ago. 2016. <http://dx.doi.org/10.1021/acs.jmedchem.5b02038>

OECD. OECD principles for the validation, for regulatory purposes, of (Quantitative) Structure-Activity Relationship models. Disponível em: <<https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>>. Acesso em: 04 jun. 2018.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2012. Disponível em: <<http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>>. Acesso em: 04 jun. 2018.

POTT, V. J. et al. Distribuição de macrófitas aquáticas numa lagoa na fazenda Nhumirim, Nhecolândia, Pantanal, MS. *Acta Botanica Brasilica*, v. 3, n. 2, p. 153–168, 1989. <http://dx.doi.org/10.1590/S0102-33061989000300015>

RAMOS; AZEVEDO. *Ecosistemas aquáticos*. [s.l: s.n.]. Campina Grande; Natal: EdUEPB; EDUFRN Editora da UFRN, 2010.

RINIKER, S.; LANDRUM, G. A. Similarity maps - A visualization strategy for molecular fingerprints and machine-learning methods. *Journal of Cheminformatics*, v. 5, n. 9, p. 1–7, 2013. <https://doi.org/10.1186/1758-2946-5-43>

RINIKER, S.; LANDRUM, G. A. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of Cheminformatics*, v. 5, n. 1, p. 26, jan. 2013b. <https://doi.org/10.1186/1758-2946-5-26>

ROGERS, D.; HAHN, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, v. 50, n. 5, p. 742–754, 2010. <https://doi.org/10.1021/ci100050t>

SIQUEIRA, J. M. et al. Estudo fitoquímico de *unonopsis lindmanii* - annonaceae, biomonitorado pelo ensaio de toxicidade sobre a *Artemia salina* leach. *Química Nova*; v. 21, n. 5, p. 557-9, 1998.

SILVA, J. M.; SANTOS, J. R. Toxicologia de agrotóxicos em ambientes aquáticos. *Oecologia Brasiliensis*, v. 11, n. 04, p. 565–573, dez. 2007. <https://doi.org/10.4257/oeco.2007.1104.08>

SOUSA, E. C. P. M. Métodos em ecotoxicologia marinha IN: NASCIMENTO, I.; SOUSA, E. C. P. M.; NIPPER, M. (Eds.) Métodos em Ecotoxicologia Marinha: Aplicações no Brasil. Editora Artes Gráficas e Indústria Ltda, São Paulo. 262p. Cap. 1:9-14. 2002.

TODESCHINI, R.; CONSONNI, V. Handbook of Molecular Descriptors. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2000. <https://doi.org/10.1002/9783527613106>

TODESCHINI, R.; CONSONNI, V. Molecular Descriptors for Chemoinformatics. 1. ed. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2009. <https://doi.org/10.1002/9783527628766>

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, v. 29, n. 6–7, p. 476–488, 6 jul. 2010. <https://doi.org/10.1002/minf.201000061>

TYLIANAKIS, J. M. The Global Plight of Pollinators. *Science*, v. 339, n. 6127, p. 1532–1533, mar. 2013. <https://doi.org/10.1126/science.1235464>

USEPA. United States Environmental Protection Agency. Guidelines for the health: risk assessment guidance for superfund (RAGS). 2002. Disponível em: <<https://www.epa.gov/risk/risk-assessment-guidance-superfund-rags-part>>. Acesso em: 12 fev. 2019.

VIERA, A. J.; GARRETT, J. M. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, v. 37, n. 5, p. 360–363, 2005.

XUE, L.; BAJORATH, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Combinatorial Chemistry & High Throughput Screening*, v. 3, n. 5, p. 363–372, 2000. <https://doi.org/10.2174/1386207003331454>

YAP, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, v. 32, n. 7, p. 1466–1474, 2011. <https://doi.org/10.1002/jcc.21707>

ZAGATTO, P. A.; BERTOLETTI, E. Ecotoxicologia aquática – princípios e aplicações. RIMA: São Paulo, 2006. 464p.

ZAGATTO, P. A.; BERTOLETTI, E. Ecotoxicologia aquática – princípios e aplicações. 2.ed. RIMA: São Paulo, 2008. 472p. ISBN: 978-85-7656-090-9