

QUIMIOINFORMÁTICA: UMA INTRODUÇÃO

Vinicius M. Alves^a, Rodolpho C. Braga^a, Eugene N. Muratov^{a,b,c} e Carolina Horta Andrade^{a,*}^aLaboratório de Planejamento de Fármacos e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, GO, 74605-220, Brasil^bLaboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA^cDepartment of Chemical Technology, Odessa National Polytechnic University, Odessa, 65000, Ukraine

Recebido em 11/06/2017; aceito em 13/09/2017; publicado na web em 08/11/2017

CHEMINFORMATICS: AN INTRODUCTION. Cheminformatics is an interdisciplinary field between chemistry and informatics, which has evolved considerably since its inception in the 1960s. Initially, the cheminformatics community dealt primarily with practical and technical aspects of chemical structure representation, manipulation, and processing, while modern research explores a new role: the exploration and interpretation of large chemical databases and the discovery of new compounds with desired activity and safety profiles. Despite the recent release of several hallmark reviews addressing methods and application of cheminformatics written in Portuguese, so far there are no scientific articles presenting cheminformatics research to the Brazilian scientific community yet. To address this gap, we aim to introduce the field of cheminformatics to both students and researchers in a simple and didactic way by narrating important historical facts and contextualizing information within the scope of various applications.

Keywords: cheminformatics; QSAR; chemical similarity; structure representation, property prediction; virtual screening.

INTRODUÇÃO

A quimioinformática é uma ciência interdisciplinar que utiliza recursos das ciências da computação e informação para resolver problemas da química.¹ O termo quimioinformática foi cunhado por Frank Brown em 1998, definindo-a como “mistura de recursos de informação para transformar dados em informação e informação em conhecimento, no intuito de tomar decisões melhores e mais rápidas na área de identificação e otimização de compostos líderes”.² Em uma definição mais abrangente, em 1999, Greg Paris, então pesquisador da companhia farmacêutica Novartis, definiu a quimioinformática como “um termo genérico que engloba a concepção, criação, organização, gestão, recuperação, análise, disseminação, visualização e utilização de informação química”.³

Apesar do nome recente, a quimioinformática não foi estabelecida ou fundada. Trata-se uma ciência que evoluiu e se consolidou ao longo de décadas.^{1,4} A revista científica mais importante da área, *Journal of Chemical Information and Modeling*, da *American Chemical Society* (ACS), existe desde 1961, chamada, na época, de *Journal of Chemical Documentation* (1961-1974) e, posteriormente, *Journal of Chemical Information and Computer Sciences* (1975-2005), chegando no nome atual em 2005 (<http://pubs.acs.org/toc/jcisd8/current>). Nesse intervalo, várias áreas que hoje compõem a quimioinformática foram se consolidando como (i) representação, visualização, manipulação e processamento de estruturas químicas, (ii) organização de bases de dados de estruturas químicas e (iii) estudos das relações quantitativas entre estrutura e atividade/propriedade (QSAR/QSPR, do inglês, *quantitative structure-activity/property relationships*).⁵

O campo da quimioinformática evoluiu bastante, passando de aspectos práticos e técnicas de representação, manipulação e processamento de estruturas químicas individuais até o seu papel primordial na atualidade: exploração de bases de dados químicas e descoberta de novos compostos com atividade e/ou propriedade desejadas. Ao se explorar bases de dados é possível extrair várias informações que

auxiliam na compreensão do comportamento de determinado grupo de compostos e é possível gerar modelos computacionais que são utilizados para prever a atividade de moléculas que carecem de dados experimentais, ou seja, dados de ensaios *in vitro* e *in vivo*.^{6,7}

Essa mudança de cenário se deve, principalmente, à grande quantidade de dados biológicos e químicos produzidos graças a evolução de ensaios de biológicos de alta vazão (HTS, do inglês, *High Throughput Screening*) e da química combinatória.⁸ A quantidade de dados cresceu tanto que, atualmente, tornou-se inviável extrair informação de bases de dados e, principalmente, transformar essa informação em conhecimento sem o uso de um computador.^{6,9} Ainda na década de 1960, quando os primeiros estudos de relação quantitativa estrutura entre atividade foram publicados, as análises eram realizadas a poucos compostos de séries congêneres e as equações deduzidas manualmente. Atualmente, modelos de QSAR/QSPR são gerados usando milhares de compostos e descritores moleculares, aplicando-se os mais variáveis algoritmos computacionais.¹⁰

É importante ressaltar que a quimioinformática está altamente relacionada e comumente converge com outras subáreas da química que trabalham com computadores,¹¹ como a química computacional,⁵ que aplica de métodos de química teórica para se calcular a estrutura e propriedades de moléculas; a modelagem molecular,¹² que usa gráficos 3D e técnicas de otimização para ajudar a compreender a natureza e ação de compostos químicos e proteínas; e o planejamento de fármacos auxiliado por computador,¹³ que diz respeito ao uso de técnicas computacionais para auxiliar na descoberta e planejamento de novas moléculas bioativas.

Nas últimas duas décadas, foram publicados artigos de revisão sobre QSAR/QSPR,^{14,15} artigos abordando metodologias de QSAR/QSPR,¹⁶⁻²⁰ e de suas aplicações,²¹⁻²⁴ além de revisões sobre modelagem molecular²⁵⁻²⁸ em língua portuguesa. Contudo, não existe até o presente momento nenhum artigo científico que apresente a quimioinformática ao público acadêmico e científico em português. Este artigo de revisão visa apresentar a quimioinformática de uma forma simples e didática. A intenção dos autores é que esse artigo sirva de material introdutório em aulas de modelagem molecular e química medicinal

*e-mail: carolina@ufg.br

e/ou computacional, e como um primeiro contato tanto para alunos de graduação quanto de pós-graduação, assim como para pesquisadores que tenham interesse em se aventurar nessa área da ciência.

REPRESENTAÇÃO DE ESTRUTURAS QUÍMICAS

Representação gráfica

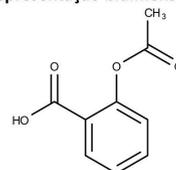
A representação gráfica de um composto químico demonstra como seus átomos estão dispostos e conectados entre si (Figura 1A, representação bidimensional e Figura 1B, representação tridimensional). Representações gráficas são úteis para que compostos sejam visualizados e compreendidos por seres humanos, mas são ineficientes para um computador capturar e processar a informação contida naquela representação.²⁹ Diversas representações computacionais de estruturas químicas já foram propostas. Aqui descrevemos as mais inovadoras e importantes. Em 1949, William Wiswesser propôs um método de representação estrutural através de linhas, denominado de *Wiswesser Line Notation* (WLN). A notação de linhas de Wiswesser foi utilizada por anos, se mostrando inovadora na triagem de subestruturas, mas esse método não produzia um código canônico (único) para cada estrutura e não era fácil de ser gerado, dando espaço à outras representações unidimensionais (1D).³⁰

Na década de 1960, Harry Morgan desenvolveu um algoritmo capaz de representar estruturas químicas de forma única.³¹ Esse algoritmo representou um importante passo na representação de estruturas, pois foi essencial para criar um registro numérico único automatizado para cada estrutura química que fosse depositada no Sistema de Registros Químicos do *Chemical Abstract Service* (CAS),³² que pertence à Sociedade Americana de Química.³³ Esse sistema representou um marco na pesquisa química, por ser uma forma simples de se identificar substâncias químicas, o que é essencial para a construção de bases de dados de compostos.

Foi na década de 1980 que David Weininger, fundador da *Daylight Chemical Information Systems, Inc.*, propôs a primeira versão do SMILES (do inglês, *Simplified Molecular-Input Line-Entry System*) (Figura 1C). Weininger também considerou a teoria dos grafos para propor seu sistema de representação de estruturas. Cada átomo é representado por seu símbolo na tabela periódica, parênteses são usados para indicar pontos de ramificação e os rótulos numéricos designam pontos de conexão de anéis. A gramática básica do SMILES também inclui informações isotópicas, configuração sobre ligações duplas e quiralidade, conhecido como SMILES isomérico.^{34,35} Desde sua criação, o SMILES foi modificado e expandido pela *Daylight* para incluir novos recursos, denominando-o de SMARTS (do inglês, *SMiles ARbitrary Target Specification*).³⁶ Posteriormente, a IUPAC (do inglês, *International Union of Pure and Applied Chemistry*) desenvolveu sua própria notação linear para representar estruturas químicas, o InChIKey (do inglês, *International Union of Pure and Applied Chemistry Key*), baseado no seu identificador InChI.³⁷

Nessa mesma época, diversos formatos CT (*Chemical Table*³⁸ ou *Connection Table*³⁹) foram desenvolvidos pela *Molecular Design Limited* (MDL), sendo o MDL MOL (ou *molfile*) e MDL SDF (ou *SDfile*) os mais utilizados (Figura 1D). Esses formatos representam as estruturas químicas como se fossem grafos e as informações são armazenadas em uma tabela. A teoria dos grafos descreve a relação de objetos em determinado conjunto através de vértices. Em arquivos CT, os átomos mais pesados que o hidrogênio correspondem aos vértices e ligações químicas às arestas.^{38,39} Existem vários *softwares* que interpretam essa informação e representam as estruturas químicas graficamente. *Softwares* como o MarvinView (<https://www.chemaxon.com/>) e PyMOL (<http://pymol.org/>), interpretam essa informação e representam as estruturas químicas graficamente.

A) Representação bidimensional



B) Representação tridimensional



C) Notações lineares

SMILES: CC(=O)Oc1ccccc1C(=O)O

SMARTS: [#6]-[#6](=O)-[#8]-[#6]-1=[#6]-[#6]=[#6]-[#6]-1-[#6]-[#8]

InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

D) Amostra de um arquivo CT (SDfile)

Mrv16a2402031713302D

```

13 13 0 0 0 0          999 v2000
1.4289 3.3000 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.4289 2.4750 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.1434 2.0625 0.0000 o 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 2.0625 0.0000 o 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 1.2375 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.4289 0.8250 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.4289 -0.0000 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 -0.4125 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.8250 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7145 1.2375 0.0000 c 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7145 2.0625 0.0000 o 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.4289 0.8250 0.0000 o 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
2 3 2 0 0 0 0
2 4 1 0 0 0 0
4 5 1 0 0 0 0
...

```

Figura 1. Métodos de representação de estruturas químicas, utilizando o ácido acetilsalicílico (Aspirina) como exemplo

Existem atualmente vários *softwares* e plataformas na web para que o usuário possa desenhar e visualizar representações de estruturas químicas, denominados editores moleculares. Um dos editores moleculares mais conhecidos, o ChemDraw⁴⁰ (<http://www.cambridgesoft.com/software/overview.aspx>), foi desenvolvido em 1985, sendo atualmente comercializado pela empresa PerkinElmer Inc. Alternativas gratuitas de livre acesso estão disponíveis e vem sendo amplamente utilizadas, como o MarvinSketch (<https://www.chemaxon.com/>) e Avogadro (<https://avogadro.cc/>). O JMSE⁴¹ (<http://peter-ertl.com/jsme/>) é um editor molecular escrito em JavaScript, que pode ser incorporado em outras ferramentas, como por exemplo, o OSRA (*Optical Structure Recognition*, <https://cactus.nci.nih.gov/cgi-bin/osra/index.cgi>), programa projetado para converter representações gráficas de estruturas químicas, como aparecem em artigos de revistas, documentos de patentes, livros didáticos, revistas comerciais, em formatos interpretáveis por um computador, como SMILES ou SDfile. Uma lista completa de editores moleculares para diversas plataformas (Windows, macOS, Linux, Java, editores online e aplicativos para celular) pode ser encontrada na página de editores moleculares do Wikipedia (https://en.wikipedia.org/wiki/Molecule_editor).

Descritores moleculares

Para que uma estrutura química seja compreendida e processada por um computador, ela precisa ser descrita em uma sequência numérica única. Descritores moleculares representam estruturas químicas incorporando muito mais informação do que os métodos anteriores. Um descritor molecular é o resultado final de um procedimento matemático e lógico que transforma informação química codificada em uma representação simbólica de uma molécula em um número útil ou é o resultado de algum experimento padronizado.⁴² Os descritores são dispostos em uma matriz (Tabela 1) ou em um vetor de *bits* (*bit vector* ou *STD logic vector*). Como pode-se observar na Tabela 1, os dados de atividade/propriedade (Y_n) são armazenados na coluna

Tabela 1. Representação genérica de uma matriz de dados contendo atividade/propriedade biológica e descritores moleculares para um estudo de QSAR/QSPR

Identificador químico	Atividade/Propriedade	Descritor 1	Descritor 2	Descritor 3	...	Descritor <i>n</i>
Composto 1	Y ₁	X ₁₁	X ₁₂	X ₁₃	...	X _{1n}
Composto 2	Y ₂	X ₂₁	X ₂₂	X ₂₃	...	X _{2n}
Composto 3	Y ₃	X ₃₁	X ₃₂	X ₃₃	...	X _{3n}
...
Composto <i>n</i>	Y _{<i>n</i>}	X _{<i>n</i>1}	X _{<i>n</i>2}	X _{<i>n</i>3}	...	X _{<i>n</i>n}

“Atividade/Propriedade” e cada descritor segue em uma nova coluna, de forma que todos os descritores (X_n) de um determinado composto estejam na mesma linha.⁴³

Diferentes tipos de descritores químicos refletem diferentes níveis de representação estrutural. Esses descritores podem ser classificados quanto à sua “dimensionalidade” em unidimensionais (1D), baseados em propriedades físico-químicas e da fórmula molecular (ex., massa molecular, refratividade molar, logP, entre outros); bidimensionais (2D), que descrevem propriedades que podem ser calculadas de uma representação 2D (ex., número de átomos, número de ligações, índices de conectividade, entre outros); e tridimensionais (3D), que dependem da conformação 3D das moléculas (ex., volume de Van der Waals, área de superfície acessível ao solvente, entre outros).⁴⁴ Outros níveis de representação, como descritores 4D, propostos por Hopfinger *et al.*,⁴⁵ constituem uma abordagem 3D que utiliza uma conformação obtida por meio de simulação de dinâmica molecular. Os descritores 5D propostos por Vedani e Dobler⁴⁶ foram uma extensão do 4D proposto por Hopfinger *et al.*,⁴⁵ adicionando liberdade conformacional, permitindo assim uma representação múltipla da topologia dos ligantes no sítio ativo. O mesmo grupo propôs em seguida descritores 6D, que consideram vários modelos de solvatação simultaneamente.⁴⁷

Outra classificação diz respeito à natureza desses descritores, podendo ser: (i) constitucionais, que são derivados da composição atômica do composto (ex., peso molecular, números de átomos e ligações); (ii) topológicos (ex., índice de conteúdo de informações de ligações); (iii) geométricos, que são derivados de coordenadas 3D (ex., volume molecular, área de superfície polar, entre outros); (iv) eletrostáticos, que são derivados das cargas parciais (ex., índices de polaridade, carga parciais, entre outros); e (v) mecânico-quânticos, que são derivados das funções de onda dos elétrons (ex., energia dos orbitais moleculares).⁴⁸ Vários programas de computador estão disponíveis para o cálculo de descritores moleculares, como DRAGON® (Talete SRL, Milan, Italy), CDK® (<http://www.rguha.net/code/java/cdkdesc.html>), CODESSA® (<http://www.semichem.com/codessa/>), RDKit (<http://www.rdkit.org/>), entre outros.

BASES DE DADOS

Atualmente, existem várias bases de dados de compostos químicos que trazem diversas informações relevantes para as mais diversas áreas da química. Como exemplo, as bases de dados ChemSpider (<http://www.chemspider.com/>) e Chemicalize (<http://chemicalize.com/>) trazem relevantes informações relacionadas à estrutura das moléculas, como nome químico, nome comercial, identificadores (e.g., número do CAS), propriedades físicas, espectro interativo, referências na literatura e fornecedores de produtos químicos. O SciFinder (<https://scifinder.cas.org/>) traz informações estruturais, reações químicas e publicações científicas e patentes relacionadas às mais de 100 milhões de estruturas químicas depositadas no CAS (<https://www.cas.org/>), base de dados de compostos químicos da

American Chemical Society, fundada em 1907.⁴⁹ A KnowItAll U (<http://www.knowitallu.com/>) provê acesso a mais de 2 milhões de espectros, incluindo infravermelho, ressonância magnética nuclear, massas, etc.

Algumas bases de dados trazem informações biológicas para estruturas químicas, como resultados de ensaios *in vitro*, *in vivo* e principalmente resultados de triagem de alta vazão (HCS/HTS). Como exemplo, cita-se a ChEMBL^{50,51} (<https://www.ebi.ac.uk/chembl/>) e PubChem^{52,53} (<http://pubchem.ncbi.nlm.nih.gov/>). Essas últimas tiveram um crescimento excepcional nos últimos anos, principalmente devido à expansão dos ensaios de triagem de alta vazão e da química combinatória, que aceleraram a quantidade de dados biológicos produzidos para moléculas pequenas.⁸ O DrugBank^{54,55} (<https://www.drugbank.ca/>) é uma base de dados de fármacos aprovados com informações mais abrangentes, como estrutura química, propriedades físico-químicas calculadas e experimentais, uso terapêutico e informações mais detalhadas, como propriedades farmacocinéticas, toxicológicas, farmacodinâmicas e sobre seu alvo molecular, quando disponíveis.

Vale ainda mencionar bases de dados de macromoléculas, como o PDB (*Protein Data Bank*, <https://www.rcsb.org/>) e o BMRDB (*Biological Magnetic Resonance Data Bank*, <http://www.bmrdb.wisc.edu/>), que são repositórios de proteínas, ácidos nucleicos e outras biomacromoléculas complexas que contribuem em estudos relacionados às ciências da saúde, planejamento de fármacos, agricultura, etc.^{56,57}

ANÁLISE DE SIMILARIDADE QUÍMICA

Frequentemente, moléculas semelhantes possuem propriedades semelhantes,^{58,59} mas isso nem sempre se aplica a todos os casos.⁶⁰ A similaridade química é um dos conceitos mais explorados em quimioinformática e em outras áreas da química, como a química medicinal⁶¹ e toxicologia.⁶² A análise de moléculas semelhantes é extremamente importante para se estabelecer relações entre estrutura e atividade ou propriedade (SAR, do inglês, *Structure-Activity Relationships*) e compreender o comportamento de determinado grupo de moléculas. Essa tarefa contribui para se encontrar erros experimentais e os denominados *cliffs* de atividade em subgrupos de moléculas. Um *cliff* de atividade é definido como um par de estruturas químicas semelhantes com atividade/propriedade muito diferentes.^{63,64} Essa análise, obviamente, pode ser feita manualmente por um bom químico, contudo, se torna inviável quando o conjunto de dados é demasiadamente grande.⁶¹

A similaridade química é calculada em um computador aplicando-se uma função de similaridade (também chamada de coeficiente de similaridade) a descritores moleculares.⁶¹ Dentre as funções de similaridade mais utilizadas, podem-se citar o coeficiente de Tanimoto⁶⁵ (também conhecido como similaridade de Jaccard),⁶⁶ e as distâncias Euclideana⁶⁷ e de Mahalanobis.⁶⁸ Qualquer tipo de descritor pode ser utilizado na análise de similaridade, mas os descritores baseados em fragmentos moleculares, principalmente os do tipo impressão

digital ou *fingerprints*, são os mais utilizados, por serem mais fáceis de se interpretar.⁶⁹

Além do auxílio na compreensão das relações entre estrutura e atividade, o uso da similaridade molecular possui várias outras aplicações. Pode-se utilizar coeficientes de similaridade para se identificar o núcleo comum de um determinado número de estruturas para se fazer uma análise de pares moleculares combinados (do inglês, *matched molecular pair analysis*, MMPA). A MMPA é uma abordagem para se identificar e comparar pares de moléculas semelhantes de um conjunto de compostos e avaliar a mudança de propriedade associada.⁷⁰ Os pares combinados são moléculas que diferem apenas por uma subestrutura particular ou um fragmento bem definido. Portanto, a MMPA pode revelar mudanças nas propriedades biológicas entre estruturas com alta similaridade, que diferem por apenas um grupo químico. Por fim, podemos citar o uso da similaridade molecular em ferramentas de busca dentro das bases de dados químicas (ver seção anterior), além de catálogos de compostos químicos comerciais, como a ChemBridge (<http://www.chembridge.com/>) e Sigma-Aldrich (<http://www.sigmaaldrich.com/>).

RELAÇÕES QUANTITATIVAS ENTRE ESTRUTURA QUÍMICA E ATIVIDADE/PROPRIEDADE (QSAR/QSPR)

Breve história e evolução

A partir da década de 1930, com base nos conhecimentos acumulados em físico-química e química orgânica, a área de relações entre estrutura e atividade ou propriedade começou a avançar.⁷¹ Em 1940, o estudo publicado por Hammett representou um marco no entendimento de propriedades moleculares de compostos orgânicos e na história do QSAR.⁷² Ele avaliou a ionização de derivados do ácido benzoico e demonstrou, pela primeira vez, uma relação linear entre a variação dos grupos substituintes e a propriedade biológica estudada. Essa relação ficou conhecida como equação de Hammett.

Quase quatro décadas se passaram quando Hansch e Fujita⁷³ consolidaram as bases dos estudos de QSAR/QSPR com um estudo inovador. Nesse estudo, os autores demonstraram que a atividade biológica poderia ser linearmente correlacionada com diferentes parâmetros físico-químicos, relacionados a efeitos hidrofóbicos, estéricos e eletrônicos. Na mesma época, Free e Wilson⁷⁴ desenvolveram uma abordagem para descrever a atividade biológica de compostos através de equações estabelecidas a partir da análise de séries congêneres. Na década de 1970, Kubinyi refinou o modelo de Hansch, desenvolvendo um modelo que descrevia a dependência não linear da atividade biológica sobre o caráter hidrofóbico.⁷⁵⁻⁷⁷

Na década de 1980, com os avanços em termos de *hardware* e *software*, o estudo das propriedades de estruturas tridimensionais (3D) tornou-se possível e então diversas abordagens de QSAR-3D foram propostas e desenvolvidas. Em 1980, Hopfinger e colaboradores desenvolveram um método baseado na forma molecular, denominado Análise da Forma Molecular (MSA, do inglês, *Molecular Shape Analysis*), o qual utiliza análise conformacional para obtenção de descritores 3D da forma molecular.⁷⁸ Em 1988, Cramer, Patterson e Bunce desenvolveram a metodologia de Análise Comparativa de Campos Moleculares (CoMFA, do inglês, *Comparative Molecular Fields Analysis*).⁷⁹ Essa técnica usa descritores 3D estéricos e eletrostáticos em campos moleculares, empregando o método dos mínimos quadrados parciais (PLS, do inglês, *partial least squares*) para estabelecer a relação com a atividade biológica. Em 1989, o CoMFA foi implementado na plataforma comercial SYBYL (<http://www.tripos.com>) e se tornou um dos métodos mais utilizados e difundidos nas últimas duas décadas.⁸⁰ No entanto, o CoMFA possui um processo de modelagem lento, além de ser dependente do alinhamento e da

conformação 3D (bioativa) das estruturas do conjunto de dados.¹⁰ Uma proposta similar ao CoMFA, mas que adicionou os efeitos de ligações de hidrogênio (doador e aceptor) e hidrofóbico foi desenvolvida pelo mesmo grupo e denominada de Análise Comparativa de Índices de Similaridade Molecular (CoMSIA, do inglês, *Comparative Molecular Similarity Index Analysis*),⁸¹ também implementado na plataforma SYBYL (<http://www.tripos.com>).

Outras metodologias de QSAR inovadoras foram desenvolvidas, como o QSAR-4D, proposto por Hopfinger e colaboradores.⁴⁵ Nesse método, os descritores são calculados para uma amostragem de conformações obtidas por meio de simulação de dinâmica molecular, reduzindo a dificuldade em encontrar a conformação bioativa. Posteriormente, extensões desse método foram propostas, como o Quasar QSAR-5D,⁴⁶ que adiciona liberdade conformacional, permitindo uma representação múltipla da topologia dos ligantes no sítio ativo, e Quasar QSAR-6D, que considera vários modelos de solvatação simultaneamente.⁴⁷

Desde o final dos anos 1990, métodos usados em QSAR/QSPR apresentaram vários avanços e crescente interesse de grupos de pesquisa e indústrias farmacêuticas.⁸² Inúmeros descritores moleculares, métodos de aprendizado de máquina e parâmetros de validação foram desenvolvidos e vêm sendo aplicados. Os estudos de QSAR/QSPR consolidaram a quimioinformática como uma ciência capaz de transformar a informação química em conhecimento. Inicialmente, modelos de QSAR/QSPR eram utilizados principalmente para otimização de compostos químicos. Atualmente, o QSAR/QSPR tem sido utilizado desde as fases de descoberta e avaliação de compostos químicos, até otimização de compostos líderes, tendendo a ser cada vez mais empregado, devido ao aprimoramento de recursos computacionais e o crescente aumento da disponibilidade de dados químicos e biológicos de alta qualidade.¹⁰

Princípios

Um modelo de QSAR é uma equação matemática que relaciona a estrutura química com a propriedade biológica. A abordagem de QSAR consiste na aplicação de vários métodos estatísticos de análise de dados para desenvolver modelos que possam prever corretamente determinada propriedade biológica de compostos baseados em sua estrutura química. Para se estabelecer essa relação, é necessário o cálculo de descritores moleculares e que a atividade biológica/propriedade tenha sido definida experimentalmente (Figura 2). O modelo de QSAR é a representação matemática final que, genericamente, pode ser definida através da Equação 1:

$$P_i = k'(D_1, D_2, \dots, D_n) \quad \text{Equação 1}$$

onde P_i é a atividade biológica ou propriedade da molécula, D_1, D_2, \dots, D_n são propriedades estruturais (descritores) calculadas (ou, em alguns casos, medidas experimentalmente) e k' é um peso definido pelo algoritmo computacional e atribuído aos descritores para se calcular as propriedades das moléculas.^{43,83}

Para se desenvolver um modelo de QSAR não é necessário, obrigatoriamente, recursos computacionais. Essa equação pode ser derivada manualmente, contudo, dado a imensa quantidade de descritores moleculares e conjuntos de dados disponíveis, esse trabalho se tornou humanamente inviável sem o auxílio de um computador. Para contornar essa dificuldade, métodos de aprendizado de máquina têm sido amplamente empregados.⁸⁴ Esses métodos estabelecem peso aos descritores, ajustando a equação que relaciona a estrutura química com a atividade biológica ou propriedade (Figura 2).⁸³

O principal objetivo de um modelo de QSAR é avaliar compostos que carecem de dados experimentais (*in vitro* e *in vivo*). Essa

abordagem possui aplicações na (i) identificação racional de novos ligantes/protótipos com atividade/propriedade desejada; (ii) otimização da atividade/propriedade; e na (iii) identificação de compostos potencialmente tóxicos.⁸³ Alguns trabalhos que demonstram a aplicação de modelos de QSAR no planejamento de fármacos e predição de toxicidade são comentados na seção *Aplicações*.

Os modelos de QSAR podem ser divididos em globais e locais. Modelos globais são gerados usando-se todo o conjunto treinamento. Nesse caso, podem existir subgrupos de moléculas que apresentam uma maior similaridade estrutural e, consequentemente, possuem características específicas. Esse subconjunto pode ser utilizado para se gerar modelos chamados de locais, ou seja, que usam um subconjunto de estruturas do conjunto de dados total baseados ou na similaridade química ou no mecanismo de ação. Apesar de modelos locais poderem contribuir com a elaboração de novas regras de relação estrutura-atividade, estudos mostram que, normalmente, não há melhorias na precisão do modelo quando se utilizam modelos de QSAR locais em comparação com modelos globais.^{85,86}

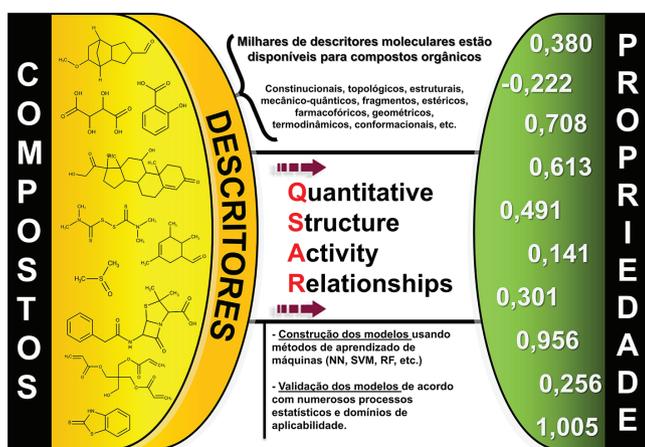


Figura 2. Esquema representando a geração de um modelo de QSAR/QSPR (modificado de ref. 83)

Aprendizado de máquina

O aprendizado de máquina consiste em métodos de inteligência artificial que dizem respeito ao estudo e construção de sistemas que podem aprender com dados. Comumente, o processo é desenvolvido em três etapas: (i) representação dos dados, (ii) otimização da hipótese, e (iii) generalização.⁸⁷ Na geração de um modelo de QSAR/QSPR, os dados são representados na forma de uma matriz com a atividade/propriedade no eixo Y e os descritores no eixo X (Tabela 1).⁸³ Uma hipótese (equação) é gerada em via de se estabelecer uma relação dos descritores com a atividade biológica. Essa relação é otimizada em um determinado número de vezes, dependendo da função utilizada e do algoritmo selecionado. Em seguida, a capacidade de generalização da hipótese é testada, avaliando-se o conjunto teste. Nesse contexto, generalizar diz respeito à habilidade da equação final em prever corretamente um exemplo/tarefa (no caso, a atividade/propriedade) não usado na construção dos modelos (conjunto teste).⁸⁷

Os métodos de aprendizado são divididos em supervisionados e não supervisionados. No primeiro caso, os algoritmos são treinados com variável Y definida na tentativa de gerar uma função matemática que generalize essa variável. Vários algoritmos de aprendizado supervisionados estão disponíveis atualmente, como *random forest (RF)*,⁸⁸ *support vector machine (SVM)*,⁸⁹ redes neurais (NN)⁹⁰ e *deep learning (DL)*.⁹¹ Os modelos gerados usando métodos supervisionados, uma vez validados, constituem um ponto de partida para a avaliação e/ou

seleção de compostos químicos que carecem de dados experimentais (ver seção *Triagem virtual à frente*).¹⁰

Os métodos não supervisionados são usados para identificar padrões nos conjuntos de dados com base apenas nos descritores, visto que a variável Y não é definida. Essa abordagem, quando aplicada aos conjuntos de dados químicos, identifica subgrupos homogêneos entre um conjunto de dados heterogêneo.⁶⁷ Alguns dos vários usos dessa abordagem incluem (i) a verificação da diversidade estrutural do conjunto de dados; (ii) a avaliação da consistência de dados experimentais; e (iii) a exploração de possíveis interferências que influenciam na atividade, contribuindo para a revelação de novas regras de SAR.^{92,93} Algoritmos comumente utilizados nessa abordagem incluem análise de componentes principais (PCA, do inglês *principal component analysis*), análise de agrupamentos hierárquicos (HCA, do inglês *hierarchical cluster analysis*) e mapas auto-organizáveis (SOM, do inglês *self-organizing maps*).⁶⁷

Boas práticas de desenvolvimento e validação

No início dos anos 2000 surgiu uma crescente preocupação com a qualidade das predições de modelos de QSAR gerados e publicados na literatura. Para garantir a qualidade dos modelos gerados, diretrizes e recomendações de boas práticas de desenvolvimento e validação de modelos de QSAR foram propostas.^{10,83,94} Em 2004, a OECD (*Organization for Economic Co-operation and Development*) publicou princípios para validação de modelos de QSAR para o uso prático na regulamentação de compostos químicos. Esses princípios são: (i) atividade biológica ou propriedade definida; (ii) algoritmo claro; (iii) domínio de aplicabilidade (DA) definido; (iv) avaliação apropriada da robustez e preditividade; (v) e interpretação mecanística, que significa encontrar relações entre os descritores e a atividade biológica ou propriedade, em via de se compreender melhor o mecanismo de ação de uma estrutura química ou aprofundar o conhecimento biológico sobre a propriedade em estudo.⁹⁵ Em consideração a esses princípios, vários elementos-chave devem ser considerados durante o desenvolvimento e validação de modelos de QSAR/QSPR e serão discutidos a seguir.^{10,83,94}

Preparo químico e biológico do conjunto de dados

Existe uma crescente preocupação sobre a irreprodutibilidade frequente de dados experimentais relatados em publicações científicas.⁹⁶ Modelos computacionais são sensíveis à qualidade dos dados utilizados e, por consequência, o preparo do conjunto de dados é indispensável para tentar evitar que erros sejam propagados em outros estudos e/ou interfiram na preditividade dos modelos. Abordagens de quimioinformática auxiliam na identificação e remoção de amostras que contêm erros experimentais ou de anotação.⁹⁷ Esse processo inclui o preparo químico e biológico dos conjuntos de dados (Figura 3).

Um fluxo de trabalho de preparo de conjunto de dados foi proposto por Fourches, Muratov e Tropsha (2010 e 2016).⁹⁸ Esse fluxo, bastante rigoroso, se inicia com o preparo químico do conjunto de dados (Figura 3, passo 1), que segue um protocolo previamente estabelecido⁹⁹ e permite a identificação e correção de erros nas estruturas químicas. Nessa etapa, misturas de componentes, compostos inorgânicos e organometálicos são removidos (quando não é possível calcular descritores para esses compostos). É feita a padronização de quimiotipos específicos, como anéis aromáticos, grupos nitro e formas tautoméricas. Contraíons são removidos e duplicatas identificadas (compostos repetidos) são analisadas e removidas (Figura 3, passo 2). A análise das duplicatas é importante pois permite avaliar a qualidade dos dados experimentais e remover estruturas químicas de registros duplicados com dados experimentais contraditórios, que afetam a qualidade dos modelos. Duplicatas presentes no conjunto

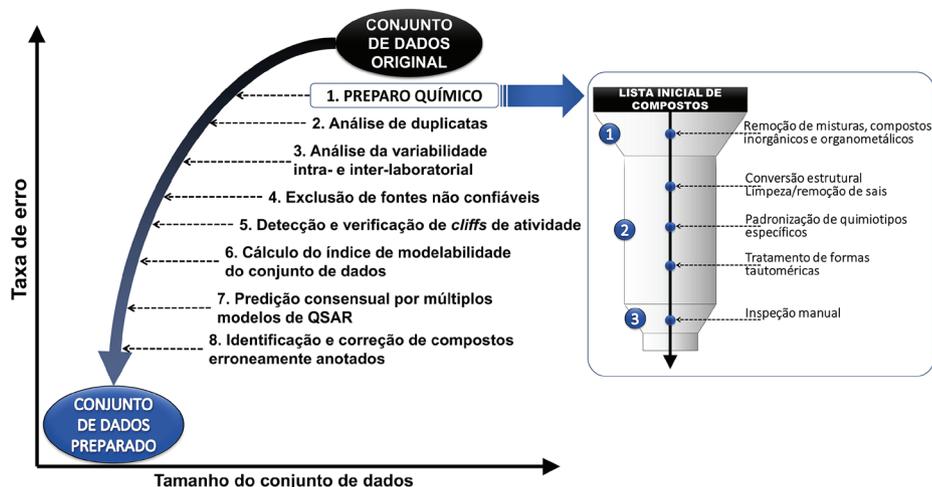


Figura 3. Fluxo geral de trabalho proposto para o preparo de conjuntos de dados (modificado de ref. 98)

teste superestimam a qualidade dos modelos. Uma inspeção manual é requerida ao final do processo para garantir que todas as estruturas estejam corretas. Em conjuntos de dados grandes, essa análise pode ser feita por amostragem.

Em seguida, realiza-se uma análise da variabilidade experimental intra- e interlaboratorial (Figura 3, passo 3) e exclui-se de fontes de dados não confiáveis, ou seja, dados com alta variação nos valores dos ensaios (passo 4), a fim de aumentar a qualidade dos dados e contribuir na tomada de decisões sobre a combinação de dados de diferentes fontes. A detecção e análise dos “cliffs” de atividade⁶³ (passo 5) e o cálculo e ajuste do índice de modelabilidade¹⁰⁰ do conjunto de dados (passo 6) estimam a viabilidade de obtenção de modelos preditivos para um determinado conjunto de dados e serve como indicadores adicionais de qualidade dos dados. A geração de modelos de QSAR de consenso (passo 7) é utilizada para tentar aumentar a confiança das predições de modelos individuais. Nessa etapa, gera-se modelos de QSAR independentes, com descritores moleculares e/ou algoritmos diferentes. O modelo de consenso é aquele que considera uma média das predições dos modelos independentes. Na última etapa (passo 8), dados biológicos incorretos podem ser identificados investigando-se compostos que tiveram atividade biológica predita muito diferente do valor experimental.

Detecção de amostras atípicas ou anômalas (“outliers”)

Amostras atípicas ou “outliers” são compostos que apresentam atividade/propriedade inesperada. Essas amostras podem resultar de erros experimentais, representar erros na estrutura química ou representarem “cliffs” de atividade. Um *cliff* de atividade representa estruturas químicas semelhantes com atividade/propriedade muito diferentes.^{63,64} Por esse motivo, amostras atípicas não devem ser removidas sem explicação, visto que sua remoção, frequentemente feita indiscriminadamente no passado para melhorar os parâmetros estatísticos do modelo, é considerada manipulação. A melhoria nos parâmetros estatísticos pode não representar melhoria real na preditividade externa do modelo. Se um composto possui um dado de atividade/propriedade dentro do intervalo do conjunto de dados e esse composto está dentro do domínio de aplicabilidade do conjunto de dados, ele não deve ser removido. Se o composto representar um *cliff* de atividade, sua má predição pelo modelo precisa ser interpretada e discutida.¹⁰

Validação dos modelos

A utilidade de um modelo de QSAR/QSPR depende de sua preditividade, ou seja, da sua capacidade em prever determinada propriedade biológica com elevada taxa de acerto. O procedimento

de validação serve para analisar a robustez e a preditividade de um modelo de QSAR. A validação é dividida em interna e externa. A validação interna é importante para determinar a robustez de um modelo, ou seja, a capacidade de reproduzir suas predições. Apesar da validação interna ser importante principalmente durante a geração dos modelos para se ajustar parâmetros do algoritmo, a real preditividade do modelo de QSAR/QSPR gerado só é avaliada usando-se um conjunto de validação externa, ou seja, que não foi utilizado para gerar, derivar, ou selecionar os modelos.^{101,102}

A validação interna normalmente é feita retirando-se um composto (*leave-one-out*) ou vários (*leave-many-out*) compostos do conjunto treinamento. Assim, um novo modelo é gerado e os compostos retirados são então preditos. Essa prática é realizada por um determinado número de vezes. O modelo final é gerado utilizando todo o conjunto treinamento. Já a validação externa é realizada separando-se parte dos compostos do conjunto modelagem para validação externa. Normalmente, separa-se entre 20% e 30% do conjunto de dados original. Um método ainda mais rigoroso de validação externa consiste no método de validação externa *n*-fold. Nesse caso, o conjunto de dados total é dividido em *n* partes. Uma das partes é utilizada como validação externa e as outras usadas para geração dos modelos. Isso é repetido *n* vezes, até que cada um dos *n* subconjuntos passe uma vez pelo conjunto de validação externa. Ao final, um consenso entre as predições individuais é realizado. A diferença entre a validação interna e externa, é que, na interna, o modelo final utiliza todos os compostos que foram utilizados para validação, ao passo que na validação externa, os compostos do conjunto de validação externa não são usados para gerar os modelos.^{103,104}

Um modelo é considerado rigorosamente validado quando ele atinge os requisitos mínimos de validação interna e externa. Diferentes métricas estatísticas foram propostas, sendo aquelas usadas para modelos contínuos diferentes das usadas para modelos categóricos. Um modelo contínuo é gerado quando a variável **Y** do conjunto de dados corresponde a uma escala numérica (p. ex., pIC50, ponto de fusão, ponto de ebulição, etc.). Um modelo categórico é gerado quando a variável **Y** é dividida em classes (p. ex., ativo/inativo, tóxico/não-tóxico, solúvel/insolúvel, etc.).

Um modelo contínuo possui boa robustez quando seu coeficiente de correlação cruzada (R^2) é maior que 0.7.¹⁰⁴ Apesar de necessária, essa métrica é insuficiente para se avaliar a preditividade externa de um modelo,¹⁰¹ por isso é necessário se avaliar o coeficiente de correlação de validação cruzada externa (Q^2), que é considerado válido quando $Q^2 \geq 0.6$.¹⁰⁴ Os modelos categóricos são validados usando-se a

acurácia balanceada (ou taxa de classificação correta),¹⁰⁵ sensibilidade e especificidade¹⁰⁶ e os valores de preditividade positiva e negativa,¹⁰⁷ sendo que todas essas métricas devem estar acima de 0.6. As métricas aqui citadas são as mais comuns em quimioinformática, contudo, existem várias outras que podem ser utilizadas em via de se garantir uma alta preditividade do modelo, como curva ROC (*receiver operating characteristic*),¹⁰⁸ diferentes fórmulas para calcular o Q^2 ,^{104,109} entre outras que podem ser encontradas na literatura.^{110,111}

Por fim, vale mencionar a técnica de aleatorização da variável **Y** (*Y-randomization*), que é recomendada para se garantir que os resultados dos modelos de QSAR não sejam provenientes do acaso. Nesse procedimento, a variável **Y** é aleatorizada e novos modelos são gerados. Caso a preditividade dos modelos com variável **Y** aleatória seja melhor que os modelos de QSAR, os modelos devem ser descartados, uma vez que os descritores moleculares não descrevem bem **Y**.¹¹²

Definição do domínio de aplicabilidade (DA)

O domínio de aplicabilidade ou domínio de aplicação é uma característica importante de qualquer modelo de QSAR.^{113,114} O DA representa o espaço químico definido pelo conjunto de moléculas do conjunto de treinamento de determinado modelo de QSAR. Predições são consideradas confiáveis quando a molécula predita se insere dentro do DA do modelo em questão. É importante ressaltar que o limite do DA é totalmente definido pelo tamanho e diversidade do conjunto utilizado para desenvolver os modelos. Quando a molécula está fora do domínio de um modelo, a predição pode ser incorreta, visto ela pode ser muito diferente e não ser completamente compreendida pelo modelo. Por esse motivo, a determinação do DA é uma etapa essencial para aceitabilidade de um modelo de QSAR.⁸³ Recentemente, um estudo¹¹⁵ demonstrou que modelos de QSAR gerados para cosméticos, fármacos e pesticidas podem ser utilizados mutuamente, ou seja, um modelo desenvolvido usando principalmente fármacos e moléculas fármaco-semelhantes pode ser usado para se avaliar cosméticos e pesticidas. Contudo, é imprescindível que antes seja verificado se a molécula a ser avaliada computacionalmente está dentro do DA daquelas usadas para gerar o modelo em questão.

TRIAGEM VIRTUAL

A triagem virtual (VS, do inglês *Virtual Screening*) compreende o processo de triagem de bibliotecas de compostos químicos através de modelos computacionais com a finalidade de avaliar e/ou selecionar compostos com propriedades desejadas.¹¹⁶ Essa é uma abordagem bem estabelecida no moderno processo de planejamento e desenvolvimento de fármacos,¹¹⁶ assim como na avaliação de potenciais compostos perigosos ao meio ambiente.⁸ A triagem virtual é uma alternativa rápida e de baixo custo para a triagem e seleção de *hits*,¹¹⁷ o que reduz o número de compostos selecionados para avaliação experimental.¹¹⁸ Habitualmente, a taxa de sucesso da triagem de alta vazão (HTS, do inglês, *High Throughput Screening*) varia entre 0,01% e 0,14%, enquanto as taxas de sucesso para a VS geralmente varia entre 1% e 40%.¹¹⁹⁻¹²¹

Os métodos computacionais usados para VS são normalmente divididos em métodos baseados no ligante (LBVS, do inglês, *Ligand-based Virtual Screening*) e na estrutura (SBVS, do inglês, *Structure-based Virtual Screening*).¹²² Abordagens baseadas no ligante usam estruturas químicas associadas a dados conhecidos para desenvolver modelos, tais como análise de similaridade,¹²³ modelos de QSAR,⁸⁴ e modelos farmacofóricos.¹²⁴ Por outro lado, os métodos baseados na estrutura utilizam a estrutura tridimensional (3D) do alvo biológico. Neste caso, as moléculas são acopladas no sítio de

ligação e classificadas com base na sua afinidade de ligação predita ou complementaridade. Modelos farmacofóricos e métodos baseados na estrutura estão fora do escopo desse artigo. Informações adicionais sobre essas técnicas podem ser encontradas na literatura.^{120,125-128}

Um método de triagem virtual que se popularizou bastante dentro da toxicologia computacional compreende o uso de alertas estruturais. Alertas estruturais¹²⁹ são subestruturas moleculares que estão associadas com determinada propriedade da molécula.¹³⁰ Comumente, alertas baseiam-se no conhecimento humano e são destinados a entender a base química do mecanismo de toxicidade ou, pelo menos, o evento molecular iniciante no caso de propriedades mais complexas.¹³¹

Os alertas estruturais são usados para sinalizar potenciais perigos e agrupar compostos em categorias para propiciar a comparação de análogos químicos por interpolação.^{132,133} Contudo, foi demonstrado que alertas estruturais são extremamente promíscuos.¹³⁴ Se usados sozinhos para prever a toxicidade, podem ser prejudiciais tanto para a avaliação da segurança quanto no planejamento de fármacos. Entretanto, os alertas estruturais, quando validados estatisticamente por modelos de QSAR, podem contribuir para o planejamento de compostos mais seguros. Em resumo, alertas estruturais devem ser usados apenas como proposta de mecanismo de ação e somente quando validados um método estatístico robusto, no caso, modelos de QSAR.¹³⁴

APLICAÇÕES DA QUIMIOINFORMÁTICA

As aplicações da quimioinformática são ilimitadas. Mesmo que seja uma área mais aplicada para a avaliação da segurança de compostos químicos, para fins ambientais e no processo de planejamento e desenvolvimento de fármacos,⁶ a quimioinformática pode ser utilizada em todas as áreas da química. Contribui para a redução do custo do desenvolvimento de novas moléculas; redução do número de animais utilizados em ensaios experimentais, e promoção da química verde. Além disso, pode aumentar a eficiência do processo de pesquisa e desenvolvimento, pois diminui resíduos gerados, visto que compostos mais improváveis de terem sucesso são descartados antes de seguirem para ensaios experimentais.^{43,135}

Um fluxograma de aplicação da quimioinformática para o planejamento de moléculas de interesse encontra-se representado na Figura 4. Dados biológicos de compostos químicos provenientes de estudos experimentais (fase 0) estão sempre sendo publicados em artigos científicos e/ou depositados em repositórios de dados (fase 1). Esses dados servem de ponto de partida para os estudos de quimioinformática. Na fase 2, os dados químicos e biológicos são compilados, preparados e integrados. Em seguida (fase 3), a análise do conjunto de dados e de similaridade é realizada e os modelos de QSAR são gerados. Os modelos de QSAR gerados são então utilizados para triar virtualmente uma biblioteca com novos compostos e/ou compostos que nunca foram avaliados experimentalmente para a propriedade em estudo (fase 4). Na fase 5, os melhores *hits* são selecionados para avaliação experimental. Compostos avaliados com propriedades indesejadas são descartados. Os compostos selecionados podem ser sintetizados ou adquiridos de empresas especializadas, como a ChemBridge (<http://www.chembridge.com/>) e Sigma-Aldrich (<https://www.sigmaaldrich.com/>). Os compostos entram novamente na fase 0 (avaliação experimental), fechando o ciclo. A avaliação experimental é a melhor forma de validação de um modelo computacional. Após essa etapa, os compostos são utilizados para enriquecer os modelos disponíveis. Essa prática aumenta o espaço químico de cobertura dos modelos e contribui para aumentar a sua preditividade.¹⁰

Modelos de QSAR têm sido empregados com sucesso na triagem virtual para a descoberta de compostos promissores. Por exemplo,

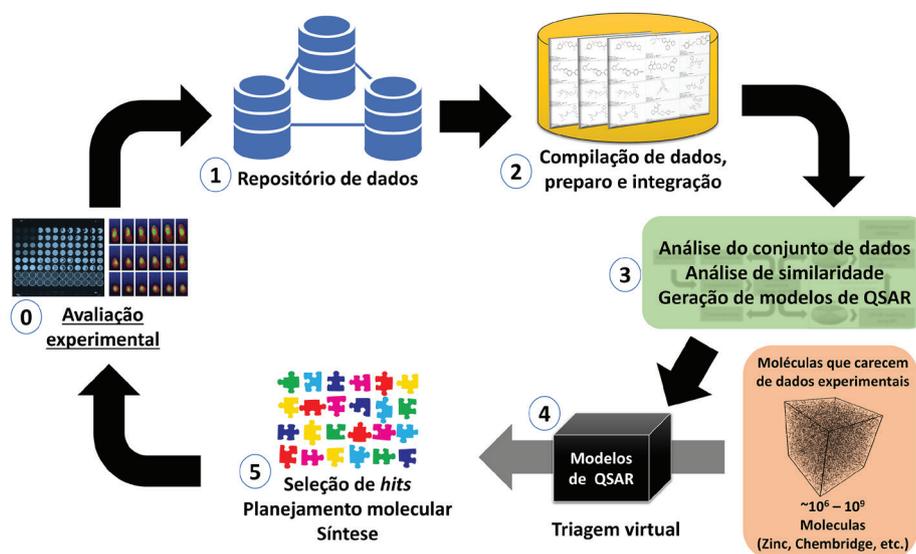


Figura 4. Aplicação da quimioinformática na avaliação, seleção e planejamento de compostos com propriedades otimizadas (modificado de ref 10).

Neves e colaboradores¹³⁶ descobriram novos compostos com atividade contra *Schistosoma mansoni* integrando técnicas de QSAR e triagem de alta vazão. Em um estudo complementar, Melo-Filho e colaboradores¹³⁷ identificaram novos esqueletos moleculares com atividade esquistosomicida. Braga e colaboradores^{138,139} desenvolveram um servidor online baseado em QSAR para a predição de toxicidade cardíaca (bloqueio da hERG). Zhang e colaboradores¹⁴⁰ descobriram 25 novos compostos com atividade antimalárica e com baixa citotoxicidade, através do emprego de modelos de QSAR. Alves e colaboradores, analisando dados experimentais e de predições de modelos de QSAR, demonstraram baixa correlação entre as propriedades de permeabilidade e sensibilização cutânea,^{141,142} descritas como correlacionadas na literatura.¹⁴³ Além disso, analisando-se dados experimentais de sensibilização cutânea de pele humana e dados do modelo animal preferencial realizado em camundongos, foi demonstrado que o modelo animal não descreve bem essa propriedade em seres humanos.¹⁴⁴

Um recente e famoso caso da aplicação do QSAR para a descoberta de fármacos consistiu na descoberta de uma nova classe de inibidores potentes e seletivos do receptor androgênico com um novo mecanismo de ação, resultando no maior acordo de licenciamento acadêmico na história do Canadá, totalizando 142 milhões de dólares.¹⁴⁵ Nesse estudo, os autores empregaram vários métodos computacionais, dentre eles, modelos de QSAR-2D, 3D e 4D, modelos farmacofóricos e acoplamento molecular. Os autores trabalharam durante todo o projeto com experimentalistas em via de otimizar os modelos e as estruturas que eram propostas.

Além dos trabalhos citados, outros estudos inovadores que compreendem a aplicação da quimioinformática para modelagem de peptídeos,^{146,147} misturas de componentes¹⁴⁸⁻¹⁵⁰ e nanopartículas¹⁵¹ também foram descritos. As aplicações dos princípios de QSAR/QSPR são várias, de forma que diversas derivações da nomenclatura têm surgido na literatura, como QSRR (*quantitative structure-(chromatographic) retention relationships*),¹⁵² QNAR (*quantitative nanostructure-activity relationships*),¹⁵³ QSTR (*quantitative structure-toxicity relationships*),¹⁵⁴ entre outros.¹⁵⁵⁻¹⁵⁷ Existem atualmente *softwares* para elucidação de estruturas químicas através de seus espectros de ressonância magnética nuclear, como ACD/Structure Elucidator Suite (http://www.acdlabs.com/products/com_iden/elucidation/struc_eluc/). Além disso, vários *softwares* como o MarvinSketch (<https://www.chemaxon.com/>) permitem cálculo de propriedades como lipofilicidade, solubilidade em água, além da predição de espectros de massa e de ressonância magnética nuclear.

Dessa maneira, observamos que métodos de quimioinformática têm sido amplamente aplicados para resolver problemas em várias áreas da química. A quimioinformática tem se consolidado em uma época em que repositórios de dados químicos e/ou biológicos com informação de livre acesso têm expandido rapidamente e sua utilidade e impacto na ciência tende a aumentar nas próximas décadas.⁶

CONSIDERAÇÕES FINAIS

A quimioinformática é uma ciência interdisciplinar que usa recursos computacionais e de tecnologia da informação para transformar informação química em conhecimento. Essa área se consolidou como uma área científica independente, evoluindo desde a década de 1960, quando lidava com aspectos práticos de representação, manipulação e processamento de estruturas químicas individuais até o seu papel primordial na atualidade: exploração de bases de dados químicos e biológicos para a descoberta de novos compostos com atividade ou propriedades desejadas.

A quimioinformática representa moléculas através de grafos armazenados em tabelas, linhas ou em descritores moleculares. Medidas de similaridade química ou relações hierárquicas entre estruturas químicas são utilizadas para se estabelecer relações entre estrutura-atividade e compreender o comportamento biológico de determinado grupo de compostos. Além disso, a quimioinformática utiliza métodos da inteligência artificial (aprendizado de máquina) para gerar modelos de QSAR que, quando devidamente validados, podem ser usados para prever a atividade ou propriedade biológica de compostos que carecem de dados experimentais.

Neste artigo de revisão, nosso principal objetivo foi apresentar, pela primeira vez em língua portuguesa, a quimioinformática de forma simples e didática, abordando aspectos históricos, conceitos, métodos e também aplicações práticas e perspectivas do progresso dessa área específica da Química. Esperamos que este material possa servir de referência para estudantes, professores e pesquisadores que estão iniciando sua jornada ou se aventurando nessa área da ciência.

AGRADECIMENTOS

Os autores agradecem ao CNPq, à CAPES e à FAPEG pelo auxílio financeiro dos projetos do grupo de pesquisa e pelas bolsas concedidas. E.M. é professor na *University of North Carolina at Chapel Hill* (EUA) e atualmente é pesquisador visitante especial

(PVE) no LabMol-UFG (CNPq # 400760/2014-2). C.H.A. é bolsista de produtividade em pesquisa do CNPq.

REFERÊNCIAS

- Gasteiger, J.; Engel, T.; *Chemoinformatics: a textbook*, Gasteiger, J.; Engel, T., eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, FRG, 2003.
- Brown, F. K.; *Annu. Rep. Med. Chem.* **1998**, *33*, 375.
- Warr, W. A.; *Extract from 218th ACS National Meeting and Exposition New Orleans, Louisiana*, August 22-26, 1999 <http://www.warr.com/warrzone2000.html>, acessada em outubro 2017.
- Hann, M.; Green, R.; *Curr. Opin. Chem. Biol.* **1999**, *3*, 379.
- Willett, P.; *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 46.
- Fourches, D.; Em *Application of Computational Techniques in Pharmacy and Medicine*; Gorb, L., Kuz'min, V., Muratov, E., eds.; Springer Netherlands: Dordrecht, 2014, cap. 16.
- Gasteiger, J.; *Molecules* **2016**, *21*, 151.
- Zhu, H.; Zhang, J.; Kim, M. T.; Boison, A.; Sedykh, A.; Moran, K.; *Chem. Res. Toxicol.* **2014**, *27*, 1643.
- Tetko, I. V.; Engkvist, O.; Chen, H.; *Future Med. Chem.* **2016**, *8*, 1801.
- Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A.; *J. Med. Chem.* **2014**, *57*, 4977.
- Wild, D.; *Introducing cheminformatics: an intensive self-guided study*, 2nd ed., McGraw-Hill Open Publishing, 2013.
- Hinchliffe, A.; *Molecular Modelling for Beginners*, Wiley: West Sussex, 2003.
- Young, D. C.; *Computational drug design: A guide for computational and medicinal chemists*, John Wiley & Sons, Inc.: Hoboken, 2009, vol. 14.
- Tavares, L. C.; *Quim. Nova* **2004**, *27*, 631.
- Gaudio, A. C.; Zandonade, E.; *Quim. Nova* **2001**, *24*, 658.
- Ferreira, M. M. C.; Montanari, C. A.; Gaudio, A. C.; *Quim. Nova* **2002**, *25*, 439.
- Arroio, A.; Honório, K. M.; da Silva, A. B. F.; *Quim. Nova* **2010**, *33*, 694.
- Montanari, M. L. C.; Montanari, C. A.; Gaudio, A. C.; *Quim. Nova* **2002**, *25*, 231.
- Cormanich, R. A.; Nunes, C. A.; Freitas, M. P.; *Quim. Nova* **2012**, *35*, 1157.
- Martins, J. P. A.; Ferreira, M. M. C.; *Quim. Nova* **2013**, *36*, 554.
- Freitas, H. F.; Paz, O. S.; Castilho, M. S.; *Quim. Nova* **2009**, *32*, 2114.
- Almeida, V. L. de; Lopes, J. C. D.; Oliveira, S. R.; Donnici, C. L.; Montanari, C. A.; *Quim. Nova* **2010**, *33*, 1482.
- Andrade, J. G.; Freitas, H. F.; Castilho, M. S.; *Quim. Nova* **2012**, *35*, 466.
- Walter, M. E.; Almeida, V. L.; Nunes, R. J.; *Quim. Nova* **2013**, *36*, 691.
- Carvalho, I.; Pupo, M. T.; Borges, Á. D. L.; Bernardes, L. S. C.; *Quim. Nova* **2003**, *26*, 428.
- Sant'Anna, C. M. R.; *Rev. Virtual Quim.* **2009**, *1*, 49.
- Barreiro, E. J.; Rodrigues, C. R.; Albuquerque, M. G.; Sant'Anna, C. M. R.; Alencastro, R. B.; *Quim. Nova* **1997**, *20*, 1.
- Andrade, C. H.; Trossini, G. H. G.; Ferreira, E. I.; *Rev. Eletronica Farm.* **2010**, *VII*, 1.
- Brecher, J.; *Pure Appl. Chem.* **2006**, *78*.
- Wiswesser, W. J.; *J. Chem. Inf. Model.* **1982**, *22*, 88.
- Morgan, H. L.; *J. Chem. Doc.* **1965**, *5*, 107.
- CAS History; <https://www.cas.org/about-cas/cas-history>, acessada em outubro 2017.
- Figueras, J.; *J. Chem. Inf. Model.* **1993**, *33*, 717.
- Weininger, D.; *J. Chem. Inf. Model.* **1988**, *28*, 31.
- Anderson, E.; Veith, G. D.; Weininger, D.; *SMILES: a line notation and computerized interpreter for chemical structures* https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=33186, acessada em outubro 2017.
- Daylight Inc.; SMARTS - A Language for Describing Molecular Patterns <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, acessada em outubro 2017.
- Pletnev, I.; Erin, A.; McNaught, A.; Blinov, K.; Tchekhovskoi, D.; Heller, S.; *J. Cheminform.* **2012**, *4*, 39.
- Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J.; *J. Chem. Inf. Model.* **1992**, *32*, 244.
- Elsevier MDL; *CTFile Formats*, Elsevier: San Leandro, 2005.
- Mills, N.; *J. Am. Chem. Soc.* **2006**, *128*, 13649.
- Bienfait, B.; Ertl, P.; *J. Cheminform.* **2013**, *5*, 24.
- Consonni, V.; Todeschini, R.; Em *Recent Advances in QSAR Studies*; Puzyn, T., Leszczynski, J., Cronin, M. T., eds.; Springer: Dordrecht, 2010, cap. 3.
- Cronin, M. T.; Em *Recent Advances in QSAR Studies*; Puzyn, T., Leszczynski, J., Cronin, M. T., eds.; Springer: Dordrecht, 2010, cap. 1.
- Xue, L.; Bajorath, J.; *Comb. Chem. High Throughput Screening* **2000**, *3*, 363.
- Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C.; *J. Am. Chem. Soc.* **1997**, *119*, 10509.
- Vedani, A.; Dobler, M.; *J. Med. Chem.* **2002**, *45*, 2139.
- Vedani, A.; Dobler, M.; Lill, M. a.; *J. Med. Chem.* **2005**, *48*, 3700.
- Todeschini, R.; Consonni, V.; *Handbook of Molecular Descriptors*, Todeschini, R., Consonni, V., eds.; Wiley: Weinheim, 2000.
- Chemical Abstracts Service; *Introduction to CAS: a division of the American Chemical Society*.
- Willighagen, E. L.; Waagmeester, A.; Spjuth, O.; Ansell, P.; Williams, A. J.; Tkachenko, V.; Hastings, J.; Chen, B.; Wild, D. J.; *J. Cheminform.* **2013**, *5*, 23.
- Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P.; *Nucleic Acids Res.* **2012**, *40*, D1100.
- Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H.; *Nucleic Acids Res.* **2012**, *40*, D400.
- Canny, S. A.; Cruz, Y.; Southern, M. R.; Griffin, P. R.; *Bioinformatics* **2012**, *28*, 140.
- Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J.; *Nucleic Acids Res.* **2006**, *34*, D668.
- Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M.; *Nucleic Acids Res.* **2008**, *36*, D901.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E.; *Nucleic Acids Res.* **2000**, *28*, 235.
- Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J.; *Nat. Struct. Biol.* **2000**, *7 Suppl*, 957.
- Johnson, M. A.; Maggiora, G. M.; *Concepts and applications of molecular similarity*, John Wiley & Sons, Inc.: New York, 1990, vol. 13.
- Heikamp, K.; Bajorath, J.; *J. Chem. Inf. Model.* **2011**, *51*, 2254.
- Kubinyi, H.; *Perspect. Drug Discovery Des.* **1998**, *9-11*, 225.
- Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J.; *J. Med. Chem.* **2014**, *57*, 3186.
- Floris, M.; Manganaro, A.; Nicolotti, O.; Medda, R.; Mangiatordi, G. F.; Benfenati, E.; *J. Cheminform.* **2014**, *6*, 39.
- Maggiora, G. M.; *J. Chem. Inf. Model.* **2006**, *46*, 1535.
- Guha, R.; Van Drie, J. H.; *J. Chem. Inf. Model.* **2008**, *48*, 646.
- Rogers, D. J.; Tanimoto, T. T.; *Science* **1960**, *132*, 1115.
- Jaccard, P.; *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *XXXVII*, 83.

67. Ferreira, M. M. C.; *Quimiometria: Conceitos, Métodos e Aplicações*, UNICAMP: Campinas, 2015.
68. Mahalanobis, P. C.; *Proc. Natl. Inst. Sci. India* **1936**, *2*, 49.
69. Willett, P.; *Drug Discov. Today* **2006**, *11*, 1046.
70. Kenny, P. W.; Sadowski, J.; Em *Cheminformatics in Drug Discovery*; Oprea, T. I., ed.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, 2005, cap. 11.
71. Hammett, L. P.; *J. Am. Chem. Soc.* **1937**, *59*, 96.
72. Hammett, L. P.; *Physical organic chemistry : reaction rates, equilibria, and mechanisms*, 1st ed., McGraw-Hill: New York, 1940.
73. Hansch, C.; Fujita, T.; *J. Am. Chem. Soc.* **1964**, *86*, 1616.
74. Free, S. M.; Wilson, J. W.; *J. Med. Chem.* **1964**, *7*, 395.
75. Kubinyi, H.; *Arzneim. Forsch.* **1977**, *27*, 750.
76. Kubinyi, H.; *Farmaco Sci.* **1979**, *34*, 248.
77. Debnath, A. K.; *Mini Rev. Med. Chem.* **2001**, *1*, 187.
78. Hopfinger, A. J.; *J. Am. Chem. Soc.* **1980**, *102*, 7196.
79. Cramer, R. D.; Patterson, D. E.; Bunce, J. D.; *J. Am. Chem. Soc.* **1988**, *110*, 5959.
80. Zhang, L.; Tsai, K.-C.; Du, L.; Fang, H.; Li, M.; Xu, W.; *Curr. Med. Chem.* **2011**, *18*, 923.
81. Klebe, G.; Abraham, U.; Mietzner, T.; *J. Med. Chem.* **1994**, *37*, 4130.
82. Cramer, R. D.; *J. Comput. Aided. Mol. Des.* **2012**, *26*, 35.
83. Tropsha, A.; *Mol. Inform.* **2010**, *29*, 476.
84. Melville, J. L.; Burke, E. K.; Hirst, J. D.; *Comb. Chem. High Throughput Screening* **2009**, *12*, 332.
85. Sheridan, R. P.; *J. Chem. Inf. Model.* **2014**, *54*, 1083.
86. Helgee, E. A.; Carlsson, L.; Boyer, S.; Norinder, U.; *J. Chem. Inf. Model.* **2010**, *50*, 677.
87. Welling, M.; *A first encounter with Machine Learning*, University of California: Irvine, 2011.
88. Breiman, L. E. O.; *Mach. Learn.* **2001**, *45*, 5.
89. Cortes, C.; Vapnik, V.; *Mach. Learn.* **1995**, *20*, 273.
90. Dreiseitl, S.; Ohno-Machado, L.; *J. Biomed. Inform.* **2002**, *35*, 352.
91. LeCun, Y.; Bengio, Y.; Hinton, G.; *Nature* **2015**, *521*, 436.
92. Downs, G. M.; Barnard, J. M.; Em *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., eds.; John Wiley & Sons, Inc.: Hoboken, 2003, vol. 18, cap. 1.
93. Mercier, D.; *Clustering large datasets*, <http://lhc.usb.vt/~mcuriel/Cursos/WC/Transfer.pdf>, acessada em outubro 2017.
94. Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E.; *SAR QSAR Environ. Res.* **2009**, *20*, 241.
95. OECD; *OECD principles for the validation, for regulatory purposes, of (Quantitative) Structure-Activity Relationship models*, <http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>, acessada em outubro 2017.
96. Frye, S. V.; Arkin, M. R.; Arrowsmith, C. H.; Conn, P. J.; Glicksman, M. A.; Hull-Ryde, E. A.; Slusher, B. S.; *Nat. Rev. Drug Discov.* **2015**, *14*, 733.
97. Fourches, D.; Muratov, E.; Tropsha, A.; *Nat. Chem. Biol.* **2015**, *11*, 535.
98. Fourches, D.; Muratov, E.; Tropsha, A.; *J. Chem. Inf. Model.* **2016**, *56*, 1243.
99. Fourches, D.; Muratov, E.; Tropsha, A.; *J. Chem. Inf. Model.* **2010**, *50*, 1189.
100. Golbraikh, A.; Muratov, E.; Fourches, D.; Tropsha, A.; *J. Chem. Inf. Model.* **2014**, *54*, 1.
101. Golbraikh, A.; Tropsha, A.; *J. Mol. Graph. Model.* **2002**, *20*, 269.
102. Gramatica, P.; *QSAR Comb. Sci.* **2007**, *26*, 694.
103. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003**, *22*, 69.
104. Chirico, N.; Gramatica, P.; *J. Chem. Inf. Model.* **2011**, *51*, 2320.
105. Wang, X. S.; Tang, H.; Golbraikh, A.; Tropsha, A.; *J. Chem. Inf. Model.* **2008**, *48*, 997.
106. Altman, D. G.; Bland, J. M.; *BMJ* **1994**, *308*, 1552.
107. Altman, D. G.; Bland, J. M.; *BMJ* **1994**, *309*, 102.
108. McClish, D. K.; *Med. Decis. Making* **1989**, *9*, 190.
109. Chirico, N.; Gramatica, P.; *J. Chem. Inf. Model.* **2012**, *52*, 2044.
110. Roy, K.; Mitra, I.; *Comb. Chem. High Throughput Screening* **2011**, *14*, 450.
111. Roy, K.; Mitra, I.; Kar, S.; Ojha, P. K.; Das, R. N.; Kabir, H.; *J. Chem. Inf. Model.* **2012**, *52*, 396.
112. Kuz'min, V. E.; Muratov, E. N.; Artemenko, A. G.; Varlamova, E. V.; Gorb, L.; Wang, J.; Leszczynski, J.; *QSAR Comb. Sci.* **2009**, *28*, 664.
113. Gadaleta, D.; Mangiatordi, G. F.; Catto, M.; Carotti, A.; Nicolotti, O.; *International Journal of Quantitative Structure-Property Relationships* **2016**, *1*, 45.
114. Mathea, M.; Klingspohn, W.; Baumann, K.; *Mol. Inform.* **2016**, *35*, 160.
115. Alves, V. M.; Muratov, E. N.; Zakharov, A.; Muratov, N. N.; Andrade, C. H.; Tropsha, A.; *Food Chem. Toxicol.* **2017**, no prelo.
116. Schneider, G.; *Nat. Rev. Drug Discov.* **2010**, *9*, 273.
117. Kessel, M.; *Nat. Biotechnol.* **2011**, *29*, 27.
118. Polgár, T.; Keseru, G. M.; *Comb. Chem. High Throughput Screening* **2011**, *14*, 889.
119. Truchon, J.-F.; Bayly, C. I.; *J. Chem. Inf. Model.* **2007**, *47*, 488.
120. Forli, S.; *Molecules* **2015**, *20*, 18732.
121. Clark, D. E.; *Expert Opin. Drug Discovery* **2008**, *3*, 841.
122. Ma, X. H.; Zhu, F.; Liu, X.; Shi, Z.; Zhang, J. X.; Yang, S. Y.; Wei, Y. Q.; Chen, Y. Z.; *Curr. Med. Chem.* **2012**, *19*, 5562.
123. Eckert, H.; Bajorath, J.; *Drug Discovery Today* **2007**, *12*, 225.
124. Yang, S.-Y.; *Drug Discovery Today* **2010**, *15*, 444.
125. Braga, R. C.; Alves, V. M.; Silva, A. C.; Nascimento, M. N.; Silva, F. C.; Liao, L. M.; Andrade, C. H.; *Curr. Top. Med. Chem.* **2014**, *14*, 1899.
126. Braga, R. C.; Andrade, C. H.; *Curr. Top. Med. Chem.* **2013**, *13*, 1127.
127. Kar, S.; Roy, K.; *Expert Opin. Drug Discovery* **2013**, *8*, 245.
128. Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J.; *J. Med. Chem.* **2010**, *53*, 8461.
129. OECD; *Report of the workshop on structural alerts for the OECD (Q) SAR application toolbox*, [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2009\)4&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2009)4&doclanguage=en), acessada em outubro 2017.
130. Blagg, J.; Em *Burger's Medicinal Chemistry and Drug Discovery*, John Wiley & Sons, Inc.: Hoboken, 2010.
131. Allen, T. E. H.; Goodman, J. M.; Gutsell, S.; Russell, P. J.; *Chem. Res. Toxicol.* **2014**, *27*, 2100.
132. Enoch, S. J.; Roberts, D. W.; Em *Chemical Toxicity Prediction: Category Formation and Read-Across*; Cronin, M., Madden, J., Enoch, S., Roberts, D., eds.; Royal Society of Chemistry, 2013, cap. 2.
133. ECHA; Agrupamento de substâncias e métodos comparativos por interpolação <https://echa.europa.eu/pt/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across>, acessada em outubro 2017.
134. Alves, V.; Muratov, E.; Capuzzi, S.; Politi, R.; Low, Y.; Braga, R.; Zakharov, A. V.; Sedykh, A.; Mokshyna, E.; Farag, S.; Andrade, C.; Kuz'min, V.; Fourches, D.; Tropsha, A.; *Green Chem.* **2016**, *18*, 4348.
135. Varnek, A.; Baskin, I. I.; *Mol. Inform.* **2011**, *30*, 20.
136. Neves, B. J.; Dantas, R. F.; Senger, M. R.; Melo-Filho, C. C.; Valente, W. C. G.; de Almeida, A. C. M.; Rezende-Neto, J. M.; Lima, E. F. C.; Paveley, R.; Furnham, N.; Muratov, E.; Kametsky, L.; Carpenter, A. E.; Braga, R. C.; Silva-Junior, F. P.; Andrade, C. H.; *J. Med. Chem.* **2016**, *59*, 7075.
137. Melo-Filho, C. C.; Dantas, R. F.; Braga, R. C.; Neves, B. J.; Senger, M. R.; Valente, W. C. G.; Rezende-Neto, J. M.; Chaves, W. T.; Muratov, E. N.; Paveley, R. A.; Furnham, N.; Kametsky, L.; Carpenter, A. E.; Silva-Junior, F. P.; Andrade, C. H.; *J. Chem. Inf. Model.* **2016**, *56*, 1357.
138. Braga, R. C.; Alves, V. M.; Silva, M. F. B.; Muratov, E.; Fourches, D.; Liao, L. M.; Tropsha, A.; Andrade, C. H.; *Mol. Inform.* **2015**, *34*, 698.

139. Braga, R. C.; Alves, V. M.; Silva, M. F. B.; Muratov, E.; Fourches, D.; Tropsha, A.; Andrade, C. H.; *Curr. Top. Med. Chem.* **2014**, *14*, 1399.
140. Zhang, L.; Fourches, D.; Sedykh, A.; Zhu, H.; Golbraikh, A.; Ekins, S.; Clark, J.; Connelly, M. C.; Sigal, M.; Hodges, D.; Guiguemde, A.; Guy, R. K.; Tropsha, A.; *J. Chem. Inf. Model.* **2013**, *53*, 475.
141. Alves, V. M.; Muratov, E. N.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; Tropsha, A.; *Toxicol. Appl. Pharmacol.* **2015**, *284*, 273.
142. Alves, V. M.; Muratov, E. N.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; Tropsha, A.; *Toxicol. Appl. Pharmacol.* **2015**, *284*, 262.
143. OECD; *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins*, http://www.oecd-ilibrary.org/environment/the-adverse-outcome-pathway-for-skin-sensitisation-initiated-by-covalent-binding-to-proteins_9789264221444-en, acessada em outubro 2017.
144. Alves, V. M.; Capuzzi, S. J.; Muratov, E.; Braga, R. C.; Thornton, T.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; Tropsha, A.; *Green Chem.* **2016**, *18*, 6501.
145. Ban, F.; Dalal, K.; Li, H.; LeBlanc, E.; Rennie, P. S.; Cherkasov, A.; *J. Chem. Inf. Model.* **2017**, *57*, 1018.
146. Lejon, T.; Strøm, M. B.; Svendsen, J. S.; *J. Pept. Sci.* **2001**, *7*, 74.
147. Lejon, T.; Stiberg, T.; Strøm, M. B.; Svendsen, J. S.; *J. Pept. Sci.* **2004**, *10*, 329.
148. Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Nikolaeva-Glomb, L.; Galabov, A. S.; Kuz'Min, V. E.; *Struct. Chem.* **2013**, *24*, 1665.
149. Oprisiu, I.; Varlamova, E.; Muratov, E.; Artemenko, A.; Marcou, G.; Polishchuk, P.; Kuz'Min, V.; Varnek, A.; *Mol. Inform.* **2012**, *31*, 491.
150. Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Kuz'min, V. E.; *Mol. Inform.* **2012**, *31*, 202.
151. Fourches, D.; Pu, D.; Li, L.; Zhou, H.; Mu, Q.; Su, G.; Yan, B.; Tropsha, A.; *Nanotoxicology* **2016**, *10*, 374.
152. Kaliszan, R.; *Chem. Rev.* **2007**, *107*, 3212.
153. Fourches, D.; Tropsha, A.; *Em Nanotoxicology: Progress toward nanomedicine*; Monteiro-Riviere, N. A., Lang Tran, C., eds.; CRC Press, 2014.
154. Roy, K.; Ghosh, G.; *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 559.
155. El-Kommos, M. E.; El-Gizawy, S. M.; Atia, N. N.; Hosny, N. M.; *J. Fluoresc.* **2015**, *25*, 1695.
156. Urbano-Cuadrado, M.; Carbó, J. J.; Maldonado, A. G.; Bo, C.; *J. Chem. Inf. Model.* *47*, 2228.
157. Garzia, A.; Villanti, A.; Tuccini, G.; *J. Pharm. Sci.* **1979**, *68*, 1081.