

BUSCA VIRTUAL DE COMPOSTOS BIOATIVOS: CONCEITOS E APLICAÇÕES

Erika Piccirillo e Antonia Tavares do Amaral*

Departamento de Química Fundamental, Instituto de Química, Universidade de São Paulo, Avenida Professor Lineu Prestes, 748, 05508-000 São Paulo – SP, Brasil

Recebido em 11/12/2017; aceito em 19/02/2018; publicado na web em 08/03/2018

VIRTUAL SCREENING OF BIOACTIVE COMPOUNDS: CONCEPTS AND APLICACIONES. The search and use of bioactive compounds for different applications have been investigated, since ancient time. Virtual screening (VS) has emerged as alternative methodological approach to the Combinatory Chemistry and High-Throughput Screening (HTS) in identifying novel drug candidates. In VS, only compounds that are selected applying different computational tools to huge virtual libraries of compounds are further tested *in vitro*. However, the effective use of VS model applications have some challenges such as the inherent complexity of the ligand-receptor interactions as well as by other factors such as ligand and receptor multiple conformations and also ligand metabolic stabilities and toxicities. Altogether these difficulties are hardly overcome using only one computational tool. Therefore, in the literature, it has been suggested to apply a sequence of different filters, such as filters that select compounds through similarity, pharmacophore and docking. In this review, we describe the advantages, limitations and examples of recent successful applications of some of these filters, including drug-like properties, structural properties, 2D similarity, pharmacophore, shape and docking filters. Moreover, we present the main steps involved in the preparation of virtual libraries of compounds that can be used in the VS.

Keywords: virtual screening; molecular fingerprints; pharmacophore; shape; docking; virtual compound libraries.

INTRODUÇÃO

O interesse pela busca e utilização de compostos bioativos, em particular daqueles usados para fins terapêuticos, já consta de relatos importantes entre as civilizações antigas, como a egípcia, a greco-romana e a chinesa.¹⁻³ Considerando o grande interesse nos compostos bioativos com uso terapêutico, estes receberão um destaque especial em nosso artigo. No entanto, gostaríamos de salientar que as abordagens aqui descritas podem também ser aplicadas na busca de compostos bioativos com outras finalidades, como, por exemplo, defensivos agrícolas.

Até a metade do século XX, os medicamentos usados pelo homem eram exclusivamente de origem natural (animal, vegetal e/ou mineral),^{4,5} sendo seus princípios ativos (fármacos), principalmente, substâncias isoladas de produtos naturais de plantas e metabólitos secundários de microrganismos.^{2,5} Dentre estes, podemos citar: a morfina (extraída da papoula, *Papaver somniferum*), a quinina (extraída de espécies de *Cinchona*) e a penicilina (extraída do fungo *Penicillium sp.*) (Figura 1 (a)). Neste período, a descoberta de fármacos era feita por “tentativa e erro”, em geral, testando-se aleatoriamente diferentes produtos naturais contra as mais diversas doenças e observando sua atividade biológica. Não raro, na literatura, encontram-se compostos descobertos ao acaso, “com base” simplesmente na dedicação/intuição/sorte do cientista, como, por exemplo, a penicilina, o clordiazepóxido, a sacarina, a fenoltaleína, entre outros.⁶ Com o avanço das Ciências Naturais, em especial da Química e da Biologia, o homem passou a estudar sistematicamente as plantas de uso medicinal.⁶ Neste contexto, as substâncias isoladas de plantas começavam a ser modificadas, de forma mais ou menos sistemática, em laboratório, visando a conseguir compostos com atividades biológicas otimizadas, ou seja, mais ativos frente ao seu alvo biológico e com menor toxicidade.⁷⁻¹⁰ Um dos primeiros exemplos de fármaco semissintético baseado em substância isolada de planta é o ácido acetilsalicílico (Aspirina®),

sintetizado por Felix Hoffmann, em 1897 (Figura 1 (a)). Este composto é uma forma menos tóxica do ácido salicílico, que é o principal metabólito da salicina (princípio ativo encontrado na casca do salgueiro, *Salix alba*). O ácido acetilsalicílico foi introduzido no mercado em 1899, sendo usado até os dias de hoje não apenas como analgésico e anti-inflamatório, mas também como um inibidor da agregação plaquetária.^{1,11}

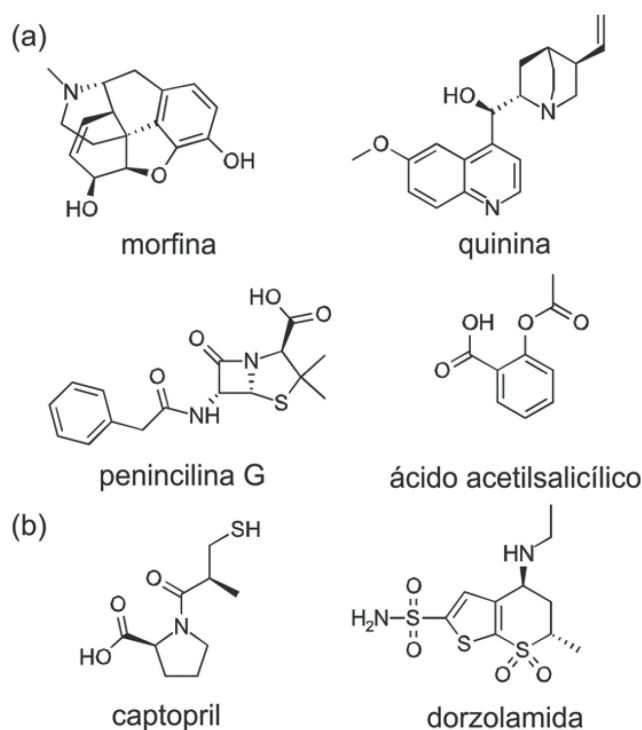


Figura 1. Representação da estrutura 2D de alguns fármacos: (a) extraídos de produtos naturais (morfina, quinina, penicilina G) e semissintético (ácido acetilsalicílico) e (b) desenvolvidos usando SBDD (captopril e dorzolamida)

*e-mail: atdamara@iq.usp.br

A descoberta do ácido acetilsalicílico marcou, de certa forma, o início da era dos fármacos sintéticos.² Assim, novos compostos eram sintetizados e testados em modelos animais e/ou preparações de órgãos. Neste período, o processo de desenvolvimento era guiado por hipóteses químicas e/ou biológicas, sendo limitantes as etapas tanto de síntese destes novos compostos quanto da avaliação de suas atividades biológicas.¹²

No início dos anos sessenta^{13,14} houve um avanço significativo no planejamento racional de compostos bioativos, em especial de fármacos, com o início dos estudos de *QSAR* (do inglês *Quantitative Structure-Activity Relationships* - Relações Quantitativas entre Estrutura química e Atividade biológica). Antes disso, vários autores já haviam tentado propor equações correlacionando a “resposta biológica/atividade fisiológica” de um determinado composto com sua estrutura química.¹⁵ No entanto, somente em 1964, com as publicações simultâneas de Hansch e Fujita¹⁴ e de Free e Wilson,¹⁶ as relações entre estrutura química e atividade biológica passaram a ser descritas de forma quantitativa, sendo este considerado o ano de nascimento do *QSAR*.¹⁵ Com o surgimento dos modelos de *QSAR*, pela primeira vez, os conceitos de físico-química orgânica foram usados para descrever de modo quantitativo a atividade biológica de um determinado composto, aplicados inicialmente para séries restritas de moléculas. Aspectos tridimensionais da interação composto-sistema biológico foram incluídos, dando origem aos modelos de *QSAR-3D*.¹⁵ Posteriormente, com o avanço na aquisição de dados tanto para os sistemas químicos como para os biológicos (vide próximos parágrafos), um grande número de informações foi gerado. Como consequência, nos últimos anos, foram identificadas ferramentas, fundamentalmente matemáticas, que permitiram decodificar este volume imenso de informações, em termos estruturais e biológicos.¹⁷ A contribuição do *QSAR* para o avanço do planejamento de fármacos é bastante reconhecida na literatura, sendo este considerado por alguns autores um dos primeiros exemplos de planejamento de fármacos com o auxílio do computador (*CADD*, do inglês *Computer-Aided Drug Design*),^{13,18} embora o termo *CADD* só tenha sido criado alguns anos depois. Uma revisão detalhada dos inúmeros estudos e aplicação dos modelos de *QSAR* em diferentes sistemas biológicos está fora do escopo deste artigo, sendo descritos em diversas publicações e revisões.¹⁹⁻²⁴

A partir dos estudos e aplicação dos modelos de *QSAR/QSAR-3D*, lentamente houve um progresso no planejamento e síntese de novos compostos bioativos, sendo, ainda, os testes biológicos a etapa limitante. Esta barreira só foi superada com a disponibilidade dos testes *in vitro* usando, por exemplo, ensaios de inibição enzimática. Os testes *in vitro* permitiram não só avaliar de forma mais rápida os analógos sintetizados, mas também aumentaram a confiabilidade dos resultados obtidos.¹² Subseqüentemente, o aparecimento da Química Combinatória²⁵ e da triagem biológica automatizada em alta escala²⁶ (*HTS*, do inglês *High-Throughput Screening*), bem como da evolução da engenharia genética, permitiram a síntese de grandes bibliotecas de compostos, que passaram a ser testados frente a seus alvos biológicos, num período de poucas semanas.²⁷ Ao mesmo tempo, o planejamento de compostos bioativos, em especial de fármacos, se beneficiou grandemente do desenvolvimento da cristalografia de raios-X de proteínas e de técnicas multidimensionais em Ressonância Magnética Nuclear (RMN).^{8,13,28-30} Estas técnicas permitiram elucidar grande número de estruturas tridimensionais (3D) de potenciais alvos biológicos, sendo grande parte delas disponível no *Protein Data Bank (PDB)*.³¹

No entanto, embora a Química Combinatória e o *HTS* tenham gerado um vasto número de compostos com diversidade estrutural, possibilitando a seleção de ligantes que apresentassem um ótimo ajuste a um determinado alvo biológico, é consenso na literatura^{10,29,32} que estas técnicas não levaram necessariamente à descoberta de novos fármacos. Alertando-se, assim, para a necessidade de se considerarem,

o mais cedo possível, outras características do ligante que o tornam um fármaco.^{10,33,34} Isto significa que, além do ótimo ajuste do ligante a um determinado alvo biológico, devem-se considerar também sua absorção, distribuição, metabolismo e eliminação e sua toxicidade, as chamadas propriedades ADME ou ADMET do ligante. Neste cenário, a inclusão de informações sobre as propriedades físico-químicas e estruturais dos ligantes e o progresso dos métodos quimiométricos desempenharam um papel fundamental.^{9,10,12,35-38}

A disponibilidade das estruturas 3D dos alvos biológicos impulsionou o surgimento de novas estratégias no planejamento de compostos bioativos, complementares às já utilizadas na época. Estas são denominadas *Structure Based Drug Design (SBDD)*, ou seja, planejamento de fármacos baseado na estrutura do alvo biológico. Por outro lado, as estratégias que não consideram a estrutura do alvo biológico passaram a ser denominadas *Ligand Based Drug Design (LBDD)*, ou seja, planejamento de fármacos baseado na estrutura do ligante. Neste artigo, os termos “alvo biológico” ou “receptor” serão usados indiscriminadamente para descrever uma macromolécula biológica na qual o composto bioativo, ou fármaco, ou ligante, deve se ligar. Normalmente esta macromolécula é uma proteína, mas também pode ser RNA ou DNA. E, ainda, o termo “ligante” será usado indiscriminadamente para se referir a qualquer molécula que se liga a uma macromolécula biológica, podendo ser um agonista, antagonista, substrato ou inibidor.

Historicamente se considera que o primeiro fármaco descoberto usando modelo estrutural da proteína foi o captopril (Figura 1 (b)), inibidor da enzima conversora de angiotensina (ECA).¹⁶ A busca de inibidores da ECA como fármacos anti-hipertensivos se iniciou em 1965, com o pesquisador brasileiro Dr. Sergio H. Ferreira, a partir do estudo dos efeitos de peptídeos extraídos de veneno de cobra no sistema renina-angiotensina-aldosterona.¹ No entanto, os resultados mais promissores só foram obtidos a partir da elucidação da estrutura 3D da carboxipeptidase A (enzima similar à ECA) complexada ao ácido L-benzil succínico. Este complexo foi usado como modelo para o sítio ativo da ECA, levando à descoberta, numa primeira etapa, do inibidor *N*-succinil-L-prolina ($IC_{50} = 330 \mu\text{mol L}^{-1}$) e, em seguida, do captopril ($IC_{50} = 23 \text{ nmol L}^{-1}$, descoberto em 1977).^{1,11} Ainda, no final dos anos noventa, foi descoberta a dorzolamida (Figura 1 (b)), um inibidor da enzima anidrase carbônica, usada no tratamento do glaucoma. A dorzolamida é o primeiro fármaco desenvolvido a partir da elucidação da estrutura 3D de seu alvo biológico, a anidrase carbônica.^{11,39}

Além do captopril e da dorzolamida, outros exemplos de fármacos descobertos empregando estratégias de *SBDD* são:¹ donepezila,⁴⁰ inibidor da acetilcolinesterase, usado no tratamento da Doença de Alzheimer; nelfinavir e amprevir,¹¹ inibidores da protease do vírus da imunodeficiência humana, usados no tratamento da síndrome da imunodeficiência adquirida; zanamivir e oseltamivir,^{41,42} inibidores da neuraminidase, antivirais usados no combate da infecção pelo vírus influenza; imatinibe,⁴³ inibidor da proteína quinase, anticancerígeno; aliscireno,⁴⁴ inibidor da renina, usado para tratar hipertensão, e sitagliptina,⁴⁵ inibidor da DPP-4 (dipeptidil peptidase 4), usada no tratamento do diabetes Tipo II. Contudo, o número de fármacos provenientes do uso de abordagens *SBDD* é, ainda, relativamente pequeno.^{1,46,47}

Mais recentemente, além de se utilizar a estrutura 3D do alvo biológico, a flexibilidade deste alvo vem sendo cada vez mais incluída nas estratégias de *SBDD*. A importância da flexibilidade conformacional para a interação ligante-receptor é amplamente aceita na literatura.⁴⁸⁻⁵⁰ No entanto, somente com os avanços computacionais obtidos nas últimas décadas, a inclusão desta flexibilidade nos protocolos de *SBDD* se tornou factível, por meio, por exemplo, de simulações de dinâmica molecular (DM) clássicas.⁴⁸⁻⁵¹

É importante ressaltar que diferentes ferramentas computacionais vêm sendo amplamente usadas na indústria farmacêutica nas etapas tanto de descoberta (fase pré-clínica) como de desenvolvimento (fase clínica). Como exemplo do uso destas ferramentas, podemos citar: na identificação de *hits*⁵² e de compostos líderes (do inglês *leads*);⁵³ no processo de otimização dos compostos líderes; na predição e otimização de suas propriedades ADMET e no desenvolvimento de formulação, entre outros.⁵⁴⁻⁵⁷

BUSCA VIRTUAL

A busca virtual, ou triagem virtual, surgiu em resposta ao insucesso obtido com o uso da Química Combinatória em conjunto com os ensaios *HTS*, já que se observou que testar aleatoriamente um grande número de compostos não resultou num avanço significativo na descoberta de novos fármacos. A busca virtual consiste em pré-selecionar os compostos com o auxílio do computador a partir de bancos de dados virtuais com um grande número de moléculas. Desta forma, apenas os compostos pré-selecionados são submetidos aos testes de atividade *in vitro*. Esta pré-seleção é feita prevendo-se, virtualmente, a atividade biológica de interesse por meio de diferentes abordagens (*LBDD* e/ou *SBDD*). Como vantagens desta estratégia têm-se que os compostos estudados não precisam, necessariamente, existir fisicamente e que neste “teste virtual”, *in silico*, não há consumo de material.²⁹

No entanto, diferentemente dos *HTS*, para se realizar a busca virtual torna-se crucial conhecer as propriedades responsáveis pela(s) interação(ões) entre o composto e o sistema biológico.^{29,58-60} Num primeiro momento, não se faz necessário considerar algumas limitações experimentais inerentes aos compostos selecionados, tais como solubilidade, formação de agregados, estabilidade, entre outras.²⁹ Contudo, a importância de evitar o quanto antes resultados falso-positivos, bem como de priorizar apenas compostos com grandes chances de sucesso na fase clínica, tem feito com que estes aspectos sejam incluídos já nas etapas iniciais de seleção. Para que os ligantes e/ou compostos selecionados possam, no futuro, se tornar fármacos é desejável considerar na busca virtual determinadas características, como sua solubilidade e suas propriedades ADMET.^{33,34,60} Sendo assim, tanto as informações disponíveis sobre a(s) interação(ões) ligante-receptor quanto sobre as propriedades do ligante são usadas na construção do modelo de busca virtual. Em seguida, este é aplicado a um banco de dados virtual de moléculas.

No entanto, a proposição de modelos de busca virtual com alta taxa de acerto é dificultada pela complexidade inerente às interações ligante-receptor em nível molecular,^{61,62} além de outros fatores, tais como: múltiplos modos de ligação do ligante,^{63,64} múltiplas conformações do ligante e do receptor,^{29,49,64,65} ligação com proteínas plasmáticas e propriedades ADMET do ligante.^{38,66,67} Tais dificuldades são raramente superadas com a utilização de uma única ferramenta computacional.⁶⁸ Como alternativa, na literatura se sugere aplicar diferentes ferramentas de modo sequencial,²⁹ como se fossem filtros de seleção. Estes filtros podem ser aplicados em qualquer sequência sem, em princípio, alterar o resultado final da seleção. Entretanto, com o objetivo de se otimizar o uso dos recursos computacionais, em geral, os filtros são aplicados em ordem crescente de complexidade do cálculo. Assim, para o banco de dados virtual de moléculas se aplicam primeiramente filtros que necessitam de baixos recursos computacionais, deixando os filtros mais complexos e elaborados para serem aplicados nas etapas finais do processo, quando o número de compostos selecionados já for significativamente reduzido. De modo geral, os filtros podem ser agrupados em duas categorias distintas: (i) filtros uni- e bidimensionais (1D e 2D), que se baseiam nas propriedades físico-químicas e na conectividade dos compostos

(filtros de propriedades “fármaco-similar” e de similaridade 2D) e (ii) filtros tridimensionais (3D), que se baseiam na distribuição espacial de características específicas, como regiões doadoras e/ouceptoras de ligação de hidrogênio ou aquelas relacionadas ao volume da molécula (filtros farmacofóricos e de “formato molecular”), e no ancoramento. As vantagens, limitações e exemplos de aplicações recentes de sucesso destes filtros de seleção serão apresentados e discutidos a seguir. O sucesso da busca virtual vai depender da qualidade tanto dos filtros de seleção quanto do banco de dados usados.⁶⁴

Além disso, cabe ressaltar que, por melhor que seja o poder de previsão do modelo proposto, este deve ser sempre submetido à validação experimental (usando-se, por exemplo, cristalografia de raios-X e/ou ensaios enzimáticos) com o objetivo de verificar sua utilidade na seleção de compostos/ligantes bioativos.^{29,46,69} De modo semelhante aos ensaios de *HTS*, os compostos selecionados por meio da busca virtual que apresentam a atividade biológica de interesse, comprovada por validação experimental, são denominados *hits* (representando resultados positivos).

FILTROS UNI- OU BIDIMENSIONAIS

Os filtros de seleção podem ser uni- ou bidimensionais, com base nas propriedades físico-químicas e/ou na conectividade dos átomos, sendo eles, por exemplo, filtros de propriedades “fármaco-similar”, filtros de características estruturais e filtros de similaridade 2D.

Filtros de propriedades “fármaco-similar”

A ideia de usar propriedades “fármaco-similar” (do inglês “*drug-like*”) como filtro de seleção na busca virtual vem da observação de que fármacos apresentam certas propriedades físico-químicas e/ou estruturais que os distinguem dos demais compostos. Dentre as características desejadas para um fármaco podemos citar: ser solúvel em água e ser, de preferência, biodisponível por via oral.

Um dos trabalhos pioneiros no sentido de identificar quais propriedades físico-químicas de um composto estão relacionadas com uma adequada biodisponibilidade oral foi desenvolvido por Lipinski e colaboradores em 1997. Neste trabalho,⁷⁰ Lipinski analisou 2.245 fármacos que apresentavam uma boa biodisponibilidade oral (absorção ou permeabilidade), utilizando quatro parâmetros físico-químicos que já haviam sido associados com medidas de permeabilidade, sendo eles: massa molecular, coeficiente de partição octanol/água calculado (usando o programa ClogP), número de átomos doadores de ligação de hidrogênio (soma dos átomos *O-H* e *N-H*) e número de átomos aceptores de ligação de hidrogênio (soma dos átomos *O* e *N*). Cerca de 90% dos compostos analisados apresentam massa molecular (MM) ≤ 500 g/mol, ClogP ≤ 5 , número de átomos doadores de ligação de hidrogênio (*HBD*, do inglês *hydrogen bond donor*) ≤ 5 e número de átomos aceptores de ligação de hidrogênio (*HBA*, do inglês *hydrogen bond acceptor*) ≤ 10 . E menos de 10% dos compostos analisados apresentam dois parâmetros fora da faixa descrita acima. Baseado nestes resultados, Lipinski sugeriu que compostos que apresentem dois ou mais parâmetros com valores fora destas faixas de corte têm grande chance de serem pouco permeáveis. A aplicação destes quatro parâmetros, acompanhadas dos valores de corte descritos acima, é chamada de “Regra dos 5” de Lipinski. Desde sua publicação em 1997, o artigo de Lipinski foi citado mais de 4.900 vezes (segundo o *Scopus*, acessado em agosto de 2017). Atualmente, a “Regra dos 5” de Lipinski é amplamente difundida na área da Química Medicinal, sendo usada rotineiramente nos protocolos de descoberta de fármacos.^{10,26,36} Uma série de outras regras para prever a biodisponibilidade oral foi derivada da regra de Lipinski, sugerindo-se, no entanto, diferentes valores de

Tabela 1. Algumas das regras derivadas da Regra dos 5 de Lipinski,⁷⁰ acompanhadas dos seus respectivos valores de corte

Regra	Valores de corte	Predição da biodisponibilidade	Referência
Regra dos 5	$MM \leq 500$; $ClogP \leq 5$, $HBD \leq 5$ e $HBA \leq 10$	Fármacos	Lipinski ⁷⁰
Regra dos 4	$MM \leq 400$; $ClogP \leq 4$, $HBD \leq 4$ e $HBA \leq 8$	Compostos líderes	Oprea ⁷¹
Regra dos 3	$MM \leq 300$; $ClogP \leq 3$, $HBD \leq 3$ e $HBA \leq 6$	Fragmentos	Ref. ⁷²
Regra dos 2	$MM \leq 200$; $ClogP \leq 2$, $HBD \leq 2$ e $HBA \leq 4$	“Blocos de construção”	Ref. ⁷³
Regra de Tice	$150 \geq MM \leq 500$; $alogP^* \leq 5$; $HBD \leq 3$; $2 \geq HBA \leq 12$ e $RTB^{**} \leq 12$	Herbicidas pós-emergenciais	Tice ⁷⁴
Regra de Tice	$150 \geq MM \leq 500$; $0 \geq alogP^* \leq 6,5$; $HBD \leq 2$; $1 \geq HBA \leq 8$ e $RTB^{**} \leq 12$	Inseticidas	Tice ⁷⁴

*Valores de logP calculados utilizando o algoritmo descrito por Ghose e Crippen.⁷⁵ **RTB é o número de ligações com livre rotação (do inglês *rotatable bonds*).

corte de acordo com os tipos de compostos analisados. A Tabela 1 mostra algumas das regras derivadas da “Regra dos 5”.

Os valores de corte descritos para os compostos líderes são menores quando comparados com os descritos para os fármacos (Tabela 1). Esta diferença se deve ao fato de que, durante a fase de otimização da atividade de um composto líder, existe uma tendência de aumento da complexidade (por exemplo: aumento da MM, aumento do número de anéis e de ligações com rotação livre) e da lipofilicidade destes compostos. Assim, para evitar que os compostos obtidos após este processo tenham baixa biodisponibilidade oral, seus valores de corte são menores. Ainda, como esperado, quanto menor o composto, menores são os correspondentes valores de corte usados. Além destes parâmetros físico-químicos, outros foram analisados e considerados como importantes para prever a biodisponibilidade oral de um fármaco, entre eles: o número de ligações com livre rotação (*RTB*, do inglês *rotatable bonds*), área de superfície polar, número de anéis aromáticos e logD calculado (coeficiente de partição no sistema octanol/água incluindo a contribuição da ionização para partição).⁷⁶⁻⁷⁸ Considerando que as abordagens apresentadas nesta revisão também podem ser usadas na descoberta de novos agroquímicos, descreve-se na literatura^{74,79} uma outra regra, derivada da de Lipinski, aplicada agora à inseticidas e à herbicidas pós-emergenciais. Esta regra foi proposta por Tice⁷⁴ considerando diferentes valores de corte para cada um dos parâmetros usados por Lipinski e incluindo ainda o número de *RTB* (Tabela 1) como parâmetro.

Além da biodisponibilidade oral, outras propriedades “farmaco-similar” relevantes podem ser incluídas nestes filtros, entre elas a solubilidade aquosa, a estabilidade metabólica e a toxicidade, sendo a solubilidade aquosa uma propriedade bastante desejada para um fármaco. No contexto da descoberta de fármacos, é mais importante garantir que o fármaco seja solúvel nas condições experimentais e/ou no sistema biológico do que determinar de forma exata seu grau de solubilidade. Além de ser importante para permitir a avaliação biológica dos fármacos, a solubilidade está associada à sua biodisponibilidade oral.⁷⁰ No entanto, a dificuldade de se prever a solubilidade aquosa de uma molécula é inerente à complexidade dos fatores envolvidos no processo de solubilização. Citam-se, entre estes, as propriedades doadoras e/ouceptoras de ligação de hidrogênio do composto e da água e o custo energético necessário para romper a estrutura cristalina no estado sólido permitindo sua solvatação. As dificuldades inerentes para se prever esta propriedade são um consenso na literatura.^{59,70,80,81} Assim, existem diversos modelos descritos na literatura⁸² baseados em descritores moleculares, bidimensionais ou tridimensionais, que variam desde simples índices até complexos modelos de *QSPR*

(*Quantitative Structure Property Relationships*). Nenhum destes modelos é capaz de prever corretamente a solubilidade aquosa, considerando todo o espaço químico conhecido. Entre os modelos de predição de solubilidade aquosa descritos na literatura, podemos citar o “*Solubility Forecast Index - SFI*”⁸⁰ e o *Soly*,⁸¹ calculado pelo programa *Volsurf+*.⁸³

O *Solubility Forecast Index - SFI* foi proposto por Hill e Young,⁸⁰ usando 36.610 compostos disponíveis na GSK. Estes foram divididos em três grupos segundo seus graus de solubilidade aquosa medidos experimentalmente: compostos com baixa solubilidade (<30 $\mu\text{mol L}^{-1}$), compostos com solubilidade intermediária (30 – 200 $\mu\text{mol L}^{-1}$) e compostos com alta solubilidade (>200 $\mu\text{mol L}^{-1}$). Os autores propuseram que a solubilidade aquosa poderia ser descrita usando os valores de logD em pH 7,4 ($clogD_{pH7,4}$, calculados pelo programa *ACDLabs*) e o número de anéis aromáticos (#Ar), dando origem ao índice *SFI* expresso pela Equação 1.

$$SFI = clogD_{pH7,4} + \#Ar \quad (1)$$

Analisando os valores de *SFI* obtidos para os compostos com baixa, com intermediária e com alta solubilidades, os autores sugeriram que compostos com valores de *SFI* >5 teriam maior probabilidade de apresentar solubilidade aquosa baixa ou intermediária.⁸⁰ Assim, este modelo não calcula o grau de solubilidade de um composto, mas o classifica segundo a probabilidade de apresentar ou não uma alta solubilidade aquosa. Embora as formas gráficas usadas para apresentar e comprovar as relações entre a solubilidade e o *SFI* tenham sido criticadas na literatura,⁸⁴ o uso de índices como o *SFI* como critério para selecionar compostos com maior probabilidade de apresentar alta solubilidade é interessante devido à simplicidade deste modelo.

Diferentemente do índice *SFI*, o modelo de solubilidade aquosa proposto por Mannhold e colaboradores⁸¹ calcula o grau de solubilidade (denominado *Soly*). Outra diferença entre estes modelos é que o último foi obtido usando o método de Mínimos Quadrados Parciais (do inglês, *Partial Least Squares - PLS*)⁸⁵ e campos de interações moleculares (explicados em detalhes a seguir). O modelo de *PLS* foi construído usando 1.028 moléculas com valores de solubilidade aquosa determinados experimentalmente, apresentando os seguintes parâmetros estatísticos: $r^2 = 0,74$; $q^2 = 0,73$ (predição interna do modelo) e erro de predição externa de ~ 0,7 em unidades de log. Devido ao erro de predição externa, os autores concluem que o modelo não deva ser usado para prever a solubilidade de compostos muito similares, podendo, no entanto, ser usado como filtro para remover compostos com valores de *Soly* abaixo de um determinado valor de corte.⁸¹ Entretanto, nenhum valor é sugerido.

Uma das críticas para a aplicação de filtros de propriedades “fármaco-similar” na construção de bancos de dados para serem usados na busca virtual é que estes reduzem o espaço químico, levando a uma possível diminuição na diversidade do banco.⁸⁶ Neste mesmo sentido, outra crítica feita à aplicação destes filtros é a recorrente observação de que alguns alvos biológicos ainda não explorados têm sítios de ligação grandes e bastante lipofílicos ou bastante polares. Nestes casos, o tamanho e as características destes sítios sugerem que a busca de *hits* e/ou compostos líderes deva ser feita usando compostos fora do “espaço químico de Lipinski”.^{78,87,88} Além disso, algumas classes de compostos (ex. antibióticos, antifúngicos e produtos naturais) não estão dentro do espaço químico definido pela regra de Lipinski, entretanto, apresentam uma boa biodisponibilidade oral.⁷⁰ Estas classes de compostos têm características estruturais que as tornam substratos de transportadores naturais. Pelo exposto, embora algumas vezes criticada, a aplicação de filtros de propriedades “fármaco-similar” durante as etapas de desenvolvimento de um fármaco pode ser, ainda, recomendada,^{10,26,36} de acordo com o sistema estudado.

Filtros de características estruturais

A aplicação desses filtros tem como objetivo remover compostos com características estruturais indesejadas em um fármaco,⁸⁹ sendo comumente denominados de “*garbage filters*”.^{12,37} Algumas destas características são associadas aos resultados falso-positivos nos ensaios *in vitro*.

A preocupação de evitar o quanto antes resultados falso-positivos no processo de descoberta de fármaco é um consenso na literatura.^{10,37,90–94} Neste sentido, esta preocupação é bem descrita pelo conceito “*fail fast, fail cheap*”.⁹⁵ A tradução literal deste conceito é “falhe rápido, falhe barato” e seu significado pode ser facilmente entendido quando se consideram tanto o tempo investido quanto os altos custos envolvidos nas fases de descoberta/desenvolvimento de fármacos (pré-clínica e clínica). Em 2014, Mullard⁹⁶ estimou que o custo para se desenvolver um fármaco pode chegar a até 2,6 bilhões de dólares. Além do custo exorbitante, este processo é bastante demorado, podendo levar de 10 a 15 anos. Assim, considerando a importância de identificar falso-positivos na etapa de descoberta (fase pré-clínica) bem como de selecionar apenas compostos com grande chance de sucesso para avançar nas etapas de desenvolvimento (fase clínica), um grande esforço vem sendo feito para entender as razões e evitar, o mais cedo possível, a ocorrência destes falso-positivos e/ou insucessos.^{10,36,90,94,97–99} Dentre elas, a presença de grupos reativos³⁷ bem como de subestruturas já reconhecidas por interferir nestes ensaios.⁹⁰ Nos próximos parágrafos, serão brevemente descritos alguns dos motivos para ocorrência de falso-positivos nos ensaios *in vitro*.

Compostos contendo grupos reativos frente a proteínas são um motivo de resultados falso-positivos, pois a atividade observada para estes compostos se deve mais à sua reatividade do que à sua capacidade de interagir com o alvo de interesse. Os exemplos mais comuns de compostos reativos atuando como falso-positivos formando ligações covalentes são aqueles que possuem em sua estrutura grupos ou centros eletrofílicos. Estes grupos são agentes alquilantes e/ou acilantes reativos frente a nucleófilos biológicos (ex. proteínas plasmáticas e glutatona). Uma lista de grupos reativos frente a nucleófilos biológicos foi descrita por Rishton¹⁰⁰ e atualizada, em 2012, por Bologna e Oprea.³⁷ A Figura 1S apresenta uma lista de alguns destes grupos reativos. E, embora estes grupos estejam presentes em algumas classes terapêuticas, Rishton defende que compostos contendo grupos reativos identificados como “ativos” devem ser avaliados com cautela para verificar se os resultados não são falso-positivos.¹⁰⁰

Em 2002, um grupo de pesquisadores da Hoffmann–La Roche¹⁰¹ analisou as características de compostos identificados como *hits* em

mais de oito ensaios *HTS* com diferentes alvos, bem como de compostos *in house* usados em pelo menos seis diferentes projetos envolvendo a descoberta de fármacos. Embora ativos, os autores salientam que as razões para a recorrente identificação destes compostos como *hits* são indesejáveis, ocorrendo por inibição inespecífica de determinados alvos (ou seja, compostos promíscuos) e/ou pela interferência com o ensaio e/ou com o método de detecção (como moléculas coloridas ou fluorescentes). Seja devido a uma atividade promíscua seja por interferirem no ensaio, os autores concluem que estes *hits* geralmente não são bons compostos líderes, sendo denominados *frequent hitters*.¹⁰¹ Embora os autores tenham conseguido separar os *frequent hitters* (479 compostos) dos não *frequent hitters* (423 compostos) usando modelos de redes neurais, eles não encontraram subestruturas comuns a todos, ou a quase todos, que pudessem ser usadas como filtros na busca virtual.

Em 2010, Baell e Holloway,⁹⁰ analisando um conjunto de 93.212 compostos da biblioteca de *HTS* do WEHIMR – *Walter and Eliza Hall Institute of Medical Research* – testados em seis diferentes campanhas de *HTS*, observaram que 2.062 (2%) deram resultados positivos em pelo menos quatro delas. Em todas estas campanhas de *HTS* foram feitos ensaios de inibição de IPP (interações proteína-proteína), usando a tecnologia *AlphaScreen*. Além dos motivos descritos acima, os autores também sugerem que o comportamento promíscuo destes compostos possa ser devido a sua capacidade de reagir com oxigênio singlete (no caso ensaios tipo *AlphaScreen*). Ainda, a partir da análise da estrutura destes *hits*, os autores sugeriram uma série de filtros estruturais agrupados de acordo com o número de compostos que apresentam a subestrutura sugerida como filtro. Por fim, os autores identificaram dez classes de compostos capazes de interferir com vários tipos de ensaios, denominados por eles como *PAINS* (do inglês, *Pan Assay INterference compoundS*) (Figura 2S). Como recomendação, sugerem evitar compostos contendo estas dez classes, embora algumas delas possam ser encontradas em fármacos já disponíveis no mercado (6,5%). Os autores salientam que, quando presentes em fármacos, estas classes também são frequentemente associadas com sua toxicidade, reatividade ou instabilidade metabólica.⁹⁰

O impacto dos resultados de Baell e Holloway⁹⁰ na área de Química Medicinal foi tão grande⁸⁹ que algumas revistas científicas de renome na área passaram a recomendar que compostos submetidos a ensaios *in vitro* fossem analisados quanto à presença de classes identificadas como *PAINS*. Pelo mesmo motivo, estas subestruturas foram disponibilizadas como filtros em vários programas usados para preparar bancos de dados virtuais de moléculas. Por sua importância, mais recentemente, dois grupos diferentes^{102,103} aplicaram e analisaram, em grande escala, estas classes identificadas como *PAINS* usando bancos de dados públicos. Entretanto, ambos mostraram que não há uma relação direta entre a presença destas classes e a frequência com que determinado composto era identificado como *hit*, como proposto. Pelo contrário, observaram que alguns dos compostos destas classes⁹⁰ foram identificados mais frequentemente como inativos do que como ativos. E, ainda, embora reconheçam a ocorrência de compostos capazes de interferir com vários tipos de ensaios,^{102,103} os autores não recomendam o uso indiscriminado dos filtros sugeridos por Baell e Holloway.⁹⁰

Considerando todos os aspectos levantados acima, a aplicação de filtros de características estruturais às bibliotecas e/ou aos bancos de dados de compostos mesmo antes que estes sejam submetidos a ensaios biológicos (*in vitro* ou *in vivo*) deve ser feita com muita cautela. Do ponto de vista dos autores desta revisão, somos da opinião que os filtros de características estruturais devem ser utilizados somente quando houver embasamento químico e/ou biológico específico que os justifique. Assim, por exemplo, num projeto de busca virtual de inibidores não covalentes de proteases, justifica-se o uso de filtros para remoção de compostos com grupamentos eletrofílicos.¹⁰⁴

Filtros de Similaridade 2D

Filtros de similaridade se baseiam no conceito de que moléculas semelhantes apresentam atividades biológicas semelhantes. No entanto, a similaridade ou dissimilaridade entre duas moléculas depende das propriedades e/ou dos critérios usados para compará-las.¹⁰⁵ Nos filtros de similaridade usando estruturas 2D, as moléculas são descritas na forma de um conjunto de caracteres denominado *fingerprint* molecular (em português: “impressão digital” molecular), representado na Figura 2.

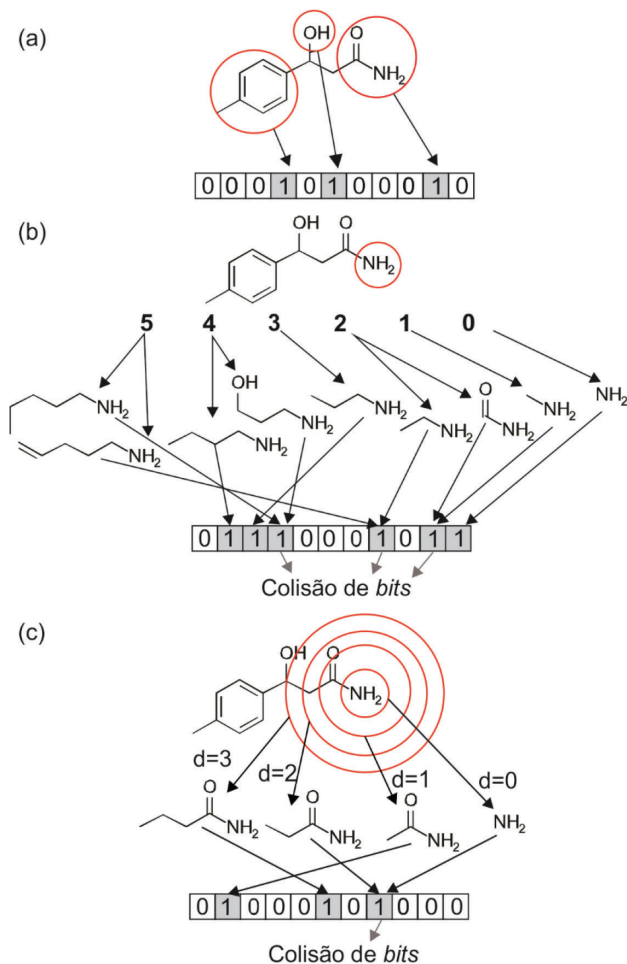


Figura 2. Representação de alguns tipos de *fingerprints* moleculares: (a) baseado em subestruturas-chave, (b) topológico ou *path-based* e (c) circular

Os *fingerprints* moleculares consistem em uma série de *bits* (valores 0 ou 1), representando a ausência (0) e a presença (1) de uma determinada subestrutura/fragmento em uma molécula (Figura 2). Estes *fingerprints* moleculares podem ser agrupados em quatro tipos:¹⁰⁶

- (1) Baseados em subestruturas-chave: os *bits* representam a presença ou a ausência de determinada subestrutura presente numa lista de subestruturas-chave (Figura 2 (a)). Este tipo de *fingerprint* só é útil para comparar moléculas que apresentam as subestruturas-chave utilizadas. O número de *bits* é igual ao número de subestruturas-chave e cada *bit* representa uma única subestrutura. Exemplos destes *fingerprints* são o *MACCS* (*Molecular ACCess System*) e o *BCI* (*Barnard Chemical Information Ltd*).
- (2) Topológicos ou *path-based*: cada molécula é analisada e dividida em fragmentos, geralmente, lineares. Em seguida, para cada fragmento se atribui um número *hash* que é usado para gerar o *fingerprint* (Figura 2 (b)). Desta forma, qualquer molécula irá

produzir um *fingerprint* e cada *bit* pode ser ocupado por mais de uma subestrutura. Exemplos destes *fingerprints* são o *The Daylight fingerprint* e o *FP2* do programa *Open Babel*.¹⁰⁷

- (3) Circulares: estes *fingerprints* são também topológicos e gerados usando número *hash*. No entanto, ao invés de fragmentos lineares, os ambientes de cada átomo são determinados por esferas (Figura 2 (c)). Por isso, estes *fingerprints* não são apropriados para serem usados como subestruturas de referência. Exemplos deste tipo de *fingerprint* são o *ECFPs* (*Extended-Connectivity Fingerprints*)¹⁰⁸ e o *FCFP* (*Functional-Class Fingerprints*).
- (4) Híbridos: estes são a combinação de mais de um tipo dos *fingerprints* descritos acima. Como exemplo temos o *UNITY 2D* e o *MP-MFP*.

A escolha do tipo de *fingerprint* molecular irá influenciar significativamente o resultado da busca virtual. Dentro do nosso conhecimento, não existe um único tipo de *fingerprint* molecular que seja sempre capaz de encontrar todos os compostos com atividade biológica. Recentemente, encontram-se na literatura publicações comparando os diferentes tipos de *fingerprints* moleculares.¹⁰⁹

Uma vez que os *fingerprints* tenham sido gerados, estes podem ser comparados usando diferentes medidas de similaridade, sendo o coeficiente de Tanimoto (Tc), representado pela Equação 2, a mais comumente empregada.^{106,110} Assim, o valor do Tc entre as moléculas A e B é calculado dividindo o número de *bits* iguais a 1 nas duas moléculas (“c”, *bits* comuns, mostrado no numerador) pelo resultado da soma do número de *bits* iguais a 1 na molécula A (“a”) e na molécula B (“b”) subtraindo-se o número de *bits* comuns às duas (“c”).

$$Tc = \frac{c}{(a+b-c)} \quad (2)$$

O valor do Tc varia de 0 a 1, significando 0 (zero) nenhuma sobreposição entre os *fingerprints* e 1 (um) uma completa sobreposição dos *fingerprints*. A magnitude do valor do Tc depende do número de *bits* presentes nos *fingerprints*. Assim, quanto maior a molécula de referência, maiores serão os valores de Tc. Num filtro de similaridade 2D, como valor de corte do Tc, sugere-se escolher aquele capaz de diferenciar compostos ativos dos inativos (denominados, geralmente, de *decoys*).¹¹¹ Na busca virtual, estes filtros podem ser usados para encontrar compostos estruturalmente semelhantes a determinado(s) composto(s) que já presente(m) a atividade biológica de interesse.¹¹² Desta forma, um dos fatores que afeta o sucesso dos filtros de similaridade 2D é a escolha da molécula de referência, pois nem todas as regiões da molécula são igualmente relevantes para a atividade biológica.

Os filtros de similaridade 2D são um dos filtros mais simples de serem utilizados na busca virtual, podendo ser aplicados mesmo quando as informações sobre a relação entre estrutura e atividade biológica são ainda bastante limitadas.¹¹² Outra aplicação da similaridade 2D é no agrupamento de compostos previamente selecionados por outros filtros, escolhendo-se um composto como representativo de cada grupo. Uma das vantagens da similaridade 2D é seu baixo custo computacional, podendo ser aplicada a bancos de dados com um grande número de compostos.¹¹² No entanto, sua principal limitação é não considerar a conformação 3D da molécula.¹¹⁰ E, ainda, não menos importante é a ocorrência de compostos similares apresentando, no entanto, atividades biológicas distintas, denominados “*activity cliffs*”.^{113–115}

FILTROS TRIDIMENSIONAIS

Os filtros tridimensionais consideram a distribuição espacial das propriedades e/ou a estrutura 3D de cada um dos compostos, ligantes

ou moléculas. Exemplos destes filtros são os farmacofóricos, os baseados no formato (do inglês “*shape*”) e o ancoramento, sendo este último o com maior custo computacional.

Filtros farmacofóricos

Os filtros farmacofóricos selecionam compostos a partir de modelos farmacofóricos. Por definição, farmacóforo é um arranjo espacial das propriedades estéricas e eletrônicas necessárias para garantir interações supramoleculares (não-covalentes) ótimas entre um ligante e seu alvo biológico específico, e, assim, ativar (ou inibir) uma resposta biológica.¹¹⁶ O farmacóforo é um conceito abstrato para considerar as características moleculares comuns em uma série de ligantes que são cruciais para sua interação com o alvo biológico (Figura 3). Assim, um farmacóforo não é uma molécula real, tampouco uma associação de grupos funcionais.¹¹⁶

Um modelo farmacofórico é um conjunto de regiões de interação (características químicas) alinhadas no espaço tridimensional. A natureza destas regiões de interação corresponde à das interações não covalentes ligante-receptor mais comumente observadas na Química Medicinal (ex. doadora e/ou acceptora de ligação de hidrogênio, eletrostática, hidrofóbica, cátion- π e π - π). Estes modelos podem ser construídos usando informações da estrutura 3D do complexo ligante-receptor (SBDD) e/ou da estrutura 3D de uma série de ligantes com atividade biológica (LBDD).^{51,111,117}

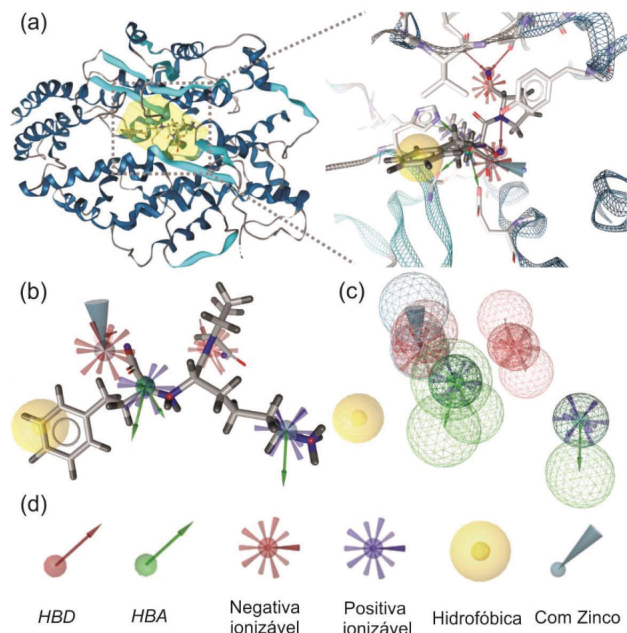


Figura 3. Representação de um modelo farmacofórico típico gerado a partir do complexo ligante-receptor: (a) reconhecimento das interações no complexo ECA-lisinopril ($K_i = 0,27 \text{ nmol L}^{-1}$)¹¹⁸ usando o programa LigandScout,¹¹⁷ (b) modelo farmacofórico sobreposto ao lisinopril, (c) modelo farmacofórico utilizado na busca virtual e (d) tipos de regiões de interação ECA-lisinopril reconhecidas pelo LigandScout,¹¹⁷ sendo HBD e HBA regiões doadora e acceptora de ligação de hidrogênio, respectivamente. Complexo ECA-lisinopril obtido diretamente do PDB (código 1O86, resolução de 2,0 Å)¹¹⁸

Quando os modelos são construídos com base na estrutura 3D do complexo ligante-receptor, a identificação das regiões de interação importantes para a atividade biológica é feita de modo intuitivo, analisando aquelas presentes no complexo (Figura 3). Por outro lado, quando os modelos são construídos com base na estrutura de um ligante, ou de uma série de ligantes, a identificação destas regiões é

mais complexa. Neste último caso, a obtenção da conformação bioativa e o alinhamento da série de compostos são etapas cruciais para a construção do modelo.¹¹⁹ Não existe um número ideal de regiões de interação para se propor um modelo farmacofórico. Acredita-se que quanto maior o número de regiões de interação menos *hits* serão selecionados.⁵¹ Modelos com muitas regiões de interação (ex. ≥ 7) selecionam um número pequeno de *hits*, que terão uma grande chance de serem similares à estrutura de referência. Por outro lado, modelos com poucas regiões de interação (ex. ≤ 3) irão selecionar *hits* inespecíficos sendo, portanto, de pouca utilidade para busca virtual.⁵¹ De modo análogo ao filtro de similaridade 2D, os filtros usando modelos farmacofóricos são susceptíveis à ocorrência dos “*activity cliffs*”. Além disso, mais especificamente para estes filtros, é importante lembrar que compostos semelhantes podem, ainda, apresentar diferentes modos de ligação.¹²⁰ Existem diversos programas disponíveis comercialmente que permitem reconhecer e gerar automaticamente modelos farmacofóricos, entre eles: CATALYST,¹²¹ FLAP,¹²² LigandScout¹¹⁷ e MOE.¹²³ Todos estes podem gerar modelos farmacofóricos partindo das estruturas tanto do receptor como do ligante.

Independentemente da forma como o modelo farmacofórico tenha sido gerado, este deve ser capaz tanto de selecionar compostos ativos como de diferenciar entre os ativos e os inativos (*decoys*). Assim, a avaliação prévia do desempenho do modelo farmacofórico gerado é bastante importante para o sucesso da busca virtual. Além de se contar o número de compostos ativos e inativos selecionados por um determinado modelo farmacofórico, na literatura são descritos outros critérios para se avaliar a qualidade dos modelos gerados.^{124–126} Os critérios mais comumente descritos são o fator de enriquecimento e a área sobre a curva (AUC, do inglês *area under the curve*), ambos obtidos a partir da curva ROC (do inglês *Receiver Operator Characteristic*, Figura 4), que mostra a fração de compostos ativos *versus* a fração de compostos inativos para um dado valor de escore (medida do ajuste do composto ao modelo farmacofórico), partindo-se dos compostos com melhor classificação.

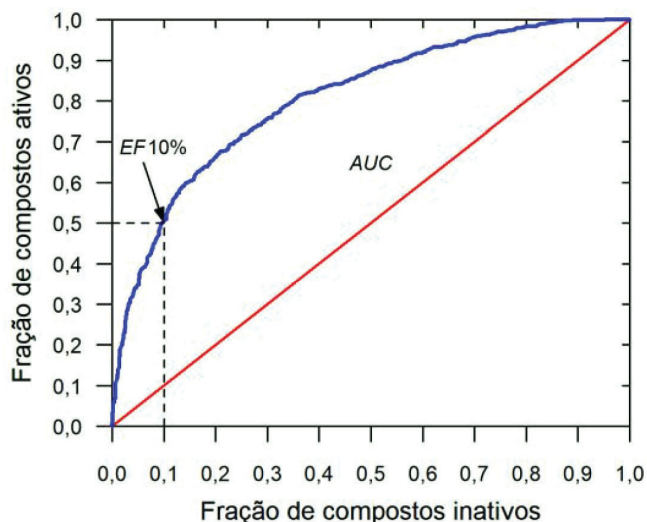


Figura 4. Representação da curva ROC típica mostrando a obtenção da AUC e do fator de enriquecimento (EF) de 10%. Nesta curva, cada ponto representa a fração de ativos/inativos selecionados para um dado valor de escore, partindo-se do maior valor de escore até o menor. A linha vermelha representa uma curva com AUC igual a 0,5 (esperada se os compostos tivessem sido selecionados aleatoriamente). Neste exemplo, os valores da AUC e do EF10% são iguais a 0,81 e a 5 (0,5/0,1), respectivamente

O fator de enriquecimento (EF) é a fração de compostos ativos (f_a) obtidos quando uma determinada porcentagem de compostos

inativos é encontrada (f_i), ou seja, $EF = f_a/f_i$. Assim, o EF depende do valor da porcentagem de compostos inativos considerado ($f_i = 1\%$, 5% ou 10%), variando de 0 a $1/f_i$. Deste modo, um modelo perfeito apresentaria um valor de $EF = 1/f_i$, enquanto um modelo que seleciona compostos ativos e inativos de modo aleatório teria um valor de $EF = 1$. O valor do AUC representa a probabilidade de um composto ativo receber um valor de escore maior que um composto inativo. Este valor varia de 0 a 1, sendo 1 o valor esperado para um modelo perfeito. Curvas com $AUC = 0,5$ indicam que os compostos ativos e inativos são selecionados de forma aleatória. Estas métricas podem ser usadas para analisar a qualidade de outros filtros usados na busca virtual que serão apresentados nos próximos itens.¹²⁴

Na literatura são encontrados vários exemplos de sucesso do uso de modelos farmacofóricos na busca virtual, permitindo identificar compostos ativos frente a diferentes alvos biológicos.¹²⁷⁻¹²⁹ Além dos compostos selecionados apresentarem atividade frente ao alvo de interesse, alguns deles têm características estruturais bastante diferentes daquelas observadas nos compostos usados para construir os modelos.

Assim, um exemplo interessante foi observado nos estudos realizados por Rechfeld e colaboradores.¹²⁹ Estes pesquisadores buscavam inibidores da interação PKC ϵ /RACK2 (interação proteína-proteína). No entanto, a ausência de informações sobre inibidores não peptídicos desta interação, associada às características do sítio ativo da PKC ϵ (grande, raso, exposto ao solvente e flexível), tornam esta tarefa bastante desafiadora. Os pesquisadores construíram um modelo farmacofórico baseado nas interações entre a PKC ϵ e um inibidor peptídico (EAVSLKPT) observadas na estrutura cristalográfica. Este modelo foi aplicado a um banco de dados contendo 300.000 moléculas, das quais foram selecionadas 468 (0,2%). A seleção final foi feita a partir da classificação de cada molécula segundo seu ajuste geométrico ao modelo farmacofórico e da análise por inspeção visual, resultando em 19 (4,0%) compostos que foram adquiridos e testados *in vitro*. Um deles apresentou atividade inibitória significativa ($IC_{50} = 25 \mu\text{mol L}^{-1}$),¹²⁹ sendo usado como estrutura de partida para construção de uma série de 19 compostos.¹²⁹ Vários compostos desta série apresentaram atividade biológica, sendo que o mais ativo apresentou $IC_{50} = 5,9 \mu\text{mol L}^{-1}$. Este estudo mostra como modelos farmacofóricos podem ser usados com sucesso na identificação de compostos bioativos, mesmo quando o alvo é considerado não *drugable* e quando as informações sobre outros inibidores são limitadas.

Modelos farmacofóricos também podem ser gerados usando apenas a estrutura do receptor (ou seja, sem qualquer ligante).¹³⁰ Neste caso, não é necessário o conhecimento prévio de ligantes. Outra vantagem dos modelos gerados deste modo é que estes não são restritos ao espaço químico ocupado pelos compostos já descritos como ligantes de um determinado alvo. Nos modelos farmacofóricos gerados deste modo, as regiões de interação são obtidas mapeando a superfície do sítio de ligação do receptor alvo. Um dos métodos mais comuns para se obterem estes modelos é por meio dos cálculos dos campos de interação molecular (do inglês *Molecular Interaction Fields - MIFs*), usando sondas moleculares de diferentes naturezas (doadora e/ou acceptora de ligação de hidrogênio, hidrofóbica e aromática). Estes campos de interação molecular podem ser calculados usando, por exemplo, o programa *GRID*.¹³¹ Durante este cálculo, o sítio de ligação do receptor alvo é colocado em uma caixa com grade regularmente espaçada. Em seguida, sondas moleculares (representando átomos ou moléculas de prova) são posicionadas nos vértices desta grade e valores de energia de interação entre a sonda e o receptor são calculados. Os valores de energia de interação calculados são usados para identificar regiões de interação favoráveis (“*hot spots*”⁵⁸) ao redor da estrutura do receptor biológico. Estes pontos de interação favoráveis podem ser usados para construir um modelo farmacofórico baseado somente na estrutura do alvo biológico.¹³⁰

Como mencionado anteriormente, além de se utilizar a estrutura 3D do alvo biológico, sua flexibilidade vem sendo cada vez mais incluída nas estratégias de *SBDD*. Uma forma de incluir a flexibilidade conformacional nos filtros farmacofóricos é gerar modelos farmacofóricos usando diferentes conformações do complexo ligante-receptor. Estas conformações podem ser obtidas experimentalmente (por cristalografia de raios-X ou por RMN) ou por computador (ex. simulações de DM clássicas).^{51,132,133}

A importância de se considerar a flexibilidade conformacional na construção dos modelos farmacofóricos foi analisada por diferentes grupos.¹³²⁻¹³⁴ Nestas análises, modelos farmacofóricos construídos usando complexos cristalográficos (“modelos rígidos”) são comparados àqueles usando complexos obtidos por simulações de DM (“modelos flexíveis”). No geral, os “modelos flexíveis” tanto obtêm um melhor desempenho quanto são capazes de selecionar compostos não identificados usando apenas os “modelos rígidos”. Uma crítica comum a esse tipo de abordagem é o pequeno número de *snapshots* (“fotografias”) usado como representante dos diferentes modos de ligação observados ao longo de toda a simulação.⁵¹ Além disso, é comum encontrar na literatura trabalhos que não utilizam qualquer critério para escolher os *snapshots* representativos das simulações (ex. selecionando o último observado na simulação). A escolha destes *snapshots* representativos é uma etapa crítica que pode impactar, significativamente, os resultados obtidos.⁴⁸

Mais recentemente, Wieder e colaboradores⁵¹ propuseram uma estratégia para incluir a flexibilidade na criação dos modelos farmacofóricos que não depende da prévia seleção de alguns *snapshots* representativos. Nesta estratégia, todos os *snapshots* obtidos durante a simulação de DM são usados para gerar modelos farmacofóricos, que são agrupados de acordo com sua semelhança. Para cada grupo, o valor da energia do ligante em cada uma das estruturas presentes no grupo é calculado. A seguir, o modelo farmacofórico que corresponde à estrutura com valor de energia médio é escolhido como representativo do correspondente grupo. Por fim, os modelos farmacofóricos representativos obtidos para um determinado alvo são usados na busca virtual. Deste modo, para cada alvo várias listas de *hits* são obtidas (uma para cada modelo representativo). Estes *hits* são agrupados gerando uma nova lista, na qual cada *hit* é contado uma única vez. A classificação dos *hits* é, então, dada pelo número de vezes que cada *hit* foi selecionado. A validade desta estratégia foi analisada usando 40 complexos ligante-receptor obtidos do *PDB*. Para cada alvo, foi usado um conjunto de compostos ativos e de compostos preditos como inativos obtidos do banco DUD-E. Como resultado, os autores observaram que o uso desta estratégia permitiu encontrar modelos melhores quando comparados aos obtidos usando somente as estruturas do *PDB*.⁵¹

Com base no exposto, a aplicação de filtros farmacofóricos independentemente de se incluir ou não a flexibilidade conformacional do receptor é uma estratégia de sucesso no caminho para a descoberta de novos compostos ativos. É importante salientar, no entanto, que este sucesso irá depender tanto do número quanto da natureza da região de interação dos modelos farmacofóricos.

Filtros baseados no formato molecular

Os filtros baseados no formato molecular (“*shape*”) selecionam moléculas com base na sua similaridade de “formato”. O formato pode ser definido como sobreposição de volumes ou de superfícies (Figura 5). A ideia de se incluir o formato como filtro vem do fato de que, se duas moléculas têm um formato similar, talvez possam apresentar propriedades similares.¹³⁵ O conceito de similaridade baseada no formato não é novo, no entanto, estes filtros passaram a ser usados rotineiramente na busca por novos compostos ativos

a partir do surgimento de ferramentas computacionais de busca em larga escala. Existem vários algoritmos de busca baseados no formato, entre eles *ROCS*¹³⁶ e *Shape Screening*¹³⁷. Entretanto, um destaque especial é dado para o programa *ROCS*,¹³⁶ pois este também é reconhecido como uma ferramenta de escolha na busca baseada no formato.¹³⁸

Além do formato da molécula, o programa *ROCS*¹³⁶ permite incluir informações sobre algumas características químicas na busca virtual (Figura 5), representando átomos ou grupos funcionais como anéis e átomos doadores ou aceptores de ligação de hidrogênio. Desta forma, diferentes tipos de medidas de similaridade podem ser calculadas com base no formato (*Shape Tanimoto*), nas características químicas (*Color Tanimoto*) e na combinação dos dois (*Combo Tanimoto*). A combinação do formato e das características químicas pode melhorar significativamente o desempenho deste filtro.

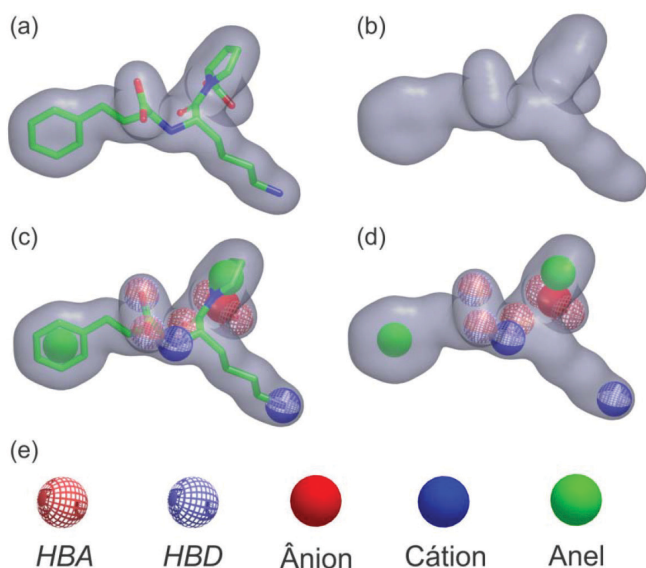


Figura 5. Representação de um modelo típico baseado no formato gerado a partir da estrutura 3D do ligante: (a) estrutura 3D do lisinopril ($K_i = 0,27 \text{ nmol L}^{-1}$)¹¹⁸ sobreposta ao modelo baseado no formato ("shape") gerado pelo programa *ROCS*,^{136,139} (b) modelo usado na busca virtual por formato, (c) estrutura 3D do lisinopril sobreposta ao modelo baseado no formato, incluindo características químicas, como gerado pelo programa *ROCS*^{136,139} e (d) modelo de formato, incluindo características químicas, usado na busca virtual e (e) alguns dos tipos de características químicas reconhecidas pelo programa *ROCS*,^{136,139} sendo HBA e HBD características aceptora e doadora de ligação de hidrogênio, respectivamente. A estrutura 3D do lisinopril foi retirada do complexo ECA-lisinopril obtido diretamente do PDB (código 1O86, resolução de 2,0 Å)¹¹⁸

De modo análogo ao descrito para os modelos farmacofóricos, para se realizar a busca por similaridade pelo formato é preciso fornecer uma determinada estrutura 3D da molécula como referência. No entanto, diferentemente dos modelos farmacofóricos baseados no ligante, este filtro parece ser menos sensível à conformação da molécula de referência, apresentando, portanto, uma vantagem em relação aos modelos farmacofóricos. Nesta abordagem, torna-se mais importante a escolha da molécula de referência do que a obtenção da conformação bioativa.¹³⁸ Ainda, como descrito para os modelos farmacofóricos, os modelos baseados no formato devem ter seu desempenho avaliado antes de serem usados como filtros na busca virtual. Esta avaliação de desempenho quanto à capacidade de diferenciar ativos de inativos pode ser feita usando as mesmas métricas já descritas para avaliar o desempenho dos modelos farmacofóricos.

Filtros de ancoramento

O ancoramento (do inglês⁸⁵ *docking*) consiste em posicionar um ligante dentro do seu sítio de ligação e, em seguida, estimar sua afinidade pela proteína alvo^{85,140} (Figura 6). A aplicação deste filtro para bibliotecas de compostos permite a seleção de acordo com seu valor de afinidade por determinada proteína alvo. Os estudos de ancoramento podem ser divididos em duas etapas: a primeira envolve a predição correta da pose do ligante (ou seja, sua conformação e orientação dentro do sítio de ligação) e a segunda envolve a predição correta da afinidade do ligante. As duas etapas são complexas e apresentam limitações.

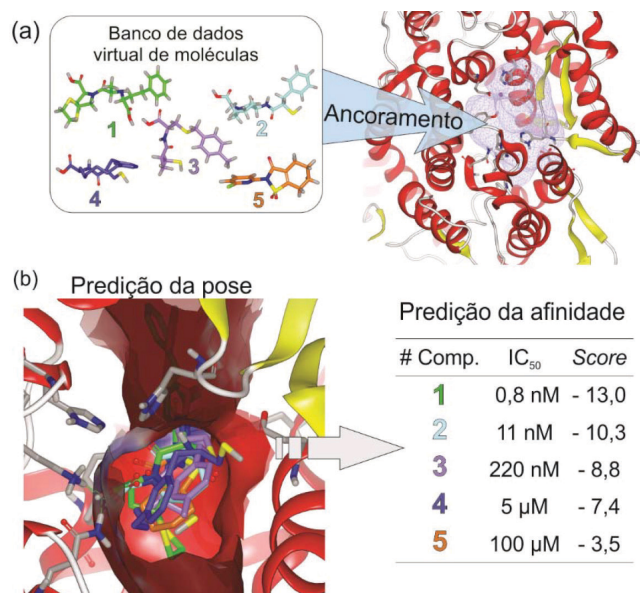


Figura 6. Representação típica do ancoramento de um banco de dados virtual de moléculas dentro do sítio ativo de determinada proteína alvo: (a) definição da região do sítio ativo considerada no ancoramento e (b) as duas etapas envolvidas no ancoramento (predições da pose e da afinidade). Ancoramento realizado usando o programa *FRED*¹⁴¹ e cinco compostos do banco de dados ChEMBL ativos frente a enzima alvo ECA. A estrutura da ECA foi obtida diretamente do PDB (código 1O86, resolução de 2,0 Å)¹¹⁸ e os compostos 1 a 5 correspondem ao ChEMBL110925, ChEMBL1560, ChEMBL1907752, ChEMBL409920 e ChEMBL2092878, respectivamente

Para que a predição da pose do ligante (ou seja, seu modo de ligação) possa ser feita de forma correta é necessário considerar todos os graus de liberdade conformacionais do ligante (translacional, rotacional e conformacional) de modo rápido e eficiente. Assim, quanto maior e mais flexível for o ligante, mais difícil deverá ser prever de forma correta seu modo de ligação. De modo geral, a varredura do espaço conformacional do ligante pode ser descrita por quatro métodos:

- (1) Baseados em campos de força. Estes usam simulações de DM ou de Monte Carlo (MC) para explorar o espaço conformacional do ligante, buscando regiões de mínimos de energia. Um dos problemas destes métodos é que os ligantes podem ficar "presos" em mínimos locais, caso os tempos de simulação sejam muito curtos. Um exemplo de programa que usa simulações de MC é o *GLIDE*.¹⁴²
- (2) Evolutivos (algoritmo genético), que fazem uma busca aleatória do espaço conformacional, diminuindo, assim, a probabilidade do ligante ficar preso em mínimos locais. Estes métodos se baseiam na teoria da evolução de Darwin. São geradas populações de cromossomos (poses do ligantes) formados por genes (ângulos diedros e vetores rotacionais e translacionais), que evoluem por

mutação de um gene e/ou cruzamento entre dois genes. Por fim, as populações com melhor escore são usadas para gerar novas populações por reprodução. Este processo se repete até que um determinado critério seja atingido. Normalmente, os programas que adotam estes métodos têm melhor desempenho quando os resultados obtidos são submetidos a uma etapa adicional de minimização para localizar o mínimo global. Exemplos de programas que usam algoritmo genético são o *AutoDock*,¹⁴³ o *AutoDock Vina*,¹⁴⁴ o *DockThor*¹⁴⁵ e o *GOLD*.¹⁴⁶

- (3) Baseados na construção incremental. Estes fazem uma busca sistemática visando a explorar todos os graus de liberdade de um determinado ligante, que é dividido em fragmentos. Em seguida, um fragmento de referência é ancorado usando a complementaridade geométrica. Este fragmento vai sendo, então, repetidamente incrementado até reconstruir totalmente o ligante. Este processo pode ser agilizado considerando apenas alguns ângulos diedros. No entanto, se na conformação bioativa o ligante apresentar ângulos torcionais não usuais, dificilmente estes métodos conseguirão encontrar a pose correta. Um exemplo de programa que usa a ancoragem incremental é o *FlexX*.¹⁴⁷
- (4) Complementaridade de forma. Estes métodos procuram por uma complementaridade estrutural entre as formas do ligante e da correspondente na cavidade da proteína. A flexibilidade do ligante é incluída usando um conjunto de conformações previamente geradas. Assim, o ancoramento é rígido, sendo apenas rotacionado e translacionado dentro do sítio ativo da proteína. Um problema deste método é que dificilmente a verdadeira conformação bioativa será encontrada dentre as previamente geradas. No entanto, uma vez que as conformações foram geradas, o ancoramento se torna mais rápido. Exemplos de programas que usam este método são o *DOCK*¹⁴⁸ e o *FRED*.¹⁴¹

O modo de ligação observado no complexo ligante-receptor obtido por cristalografia de raios-X é considerado, dentro do erro experimental, como uma das possíveis conformações bioativas do ligante. Assim, o desempenho dos programas de ancoramento quanto à obtenção da pose correta é, normalmente, avaliado observando-se quanto os resultados obtidos reproduzem estes dados experimentais. Em geral, todos os programas de ancoramento, usando qualquer um dos quatro métodos descritos acima, são capazes de reproduzir satisfatoriamente (com valores de $RMSD^{149} \leq 2\text{Å}$) as poses dos ligantes observadas nos complexos obtidos por cristalografia.¹⁵⁰ Este tipo de avaliação de desempenho é chamado de *re-docking*.¹⁵¹

A segunda etapa do ancoramento é prever e classificar a afinidade de diferentes poses de um mesmo ligante, e/ou de diferentes ligantes, por determinada proteína alvo. Esta etapa é fundamental para o sucesso do ancoramento, pois o programa deve ser capaz de diferenciar as poses corretas das incorretas, bem como os compostos ativos dos inativos. Métodos de cálculo de energia de *Gibbs* usando mecânica estatística foram desenvolvidos para modelar as interações proteína-ligante e, assim, prever a afinidade desta ligação. Entretanto, estes métodos ainda são computacionalmente muito caros para serem usados como forma de avaliação de um grande número de complexos ligante-receptor. Como alternativa, os programas de ancoramento preveem a afinidade ligante-receptor por meio de funções de ranqueamento. Estas funções fazem várias suposições e simplificações não considerando completamente todos os aspectos envolvidos no reconhecimento ligante-receptor, como, por exemplo, as contribuições entrópicas e a polarizabilidade. De modo geral, as funções de ranqueamento podem ser agrupadas em três classes:

- (1) Baseadas em campos de força molecular: calculam valores de energia da interação ligante-receptor usando campos de força moleculares da mecânica clássica. A principal desvantagem

destas funções é não considerar os fatores entrópicos associados à interação ligante-receptor. Assim, deve-se ter cuidado para não superestimar as altas afinidades calculadas para moléculas grandes e/ou polares. Um exemplo deste tipo de função de ranqueamento é a *Goldscore*.¹⁵²

- (2) Empíricas: calculam o valor de energia da interação ligante-receptor usando uma equação complexa com vários termos, que descreve propriedades reconhecidas como importantes para a interação ligante-receptor. Séries de treinamento e métodos de regressão multivariada são usados para otimizar os valores dos coeficientes que descrevem a contribuição de cada termo para a equação. Assim, é esperado que as equações obtidas para um conjunto de determinadas proteínas só possam ser aplicadas para proteínas similares às presentes na série de treinamento. Exemplos de programa que usam funções de ranqueamento empíricas são o *FRED*¹⁴¹ e o *GOLD*.¹⁴⁶
- (3) Baseadas no conhecimento de potenciais médios de força: calculam os valores de energia da interação ligante-receptor usando a soma das energias livre de interação de todos os pares de átomos ligante-receptor. Estas energias são calculadas considerando os valores de distância entre dois pares de átomos observados em complexos ligante-receptor obtidos por cristalografia. As vantagens destas funções é que não são necessárias séries de testes e que os termos entrópicos e relacionados à dessolvatação são incluídos implicitamente. No entanto, a definição arbitrária dos tipos de átomos envolvidos na interação é um dos fatores que limita a aplicabilidade destas funções de ranqueamento. Um exemplo de programa que usa estas funções é o *FlexX*.¹⁴⁷

Estudos comparativos mostram que todas as funções de ranqueamento têm enorme dificuldade para identificar a pose mais correta dentre todas as poses geradas, sendo bastante importante a inspeção visual dos resultados obtidos. Além disso, outra limitação das funções de ranqueamento é a predição absoluta da afinidade do ligante pelo seu alvo. Melhores resultados são obtidos quando se analisa a capacidade das funções classificarem de forma mais ou menos correta a atividade relativa de um conjunto de ligantes. No entanto, o desempenho das funções de ranqueamento está longe de ser ideal e parece ser bastante dependente do alvo molecular usado.^{29,46,140,150,153,154} A avaliação do desempenho destas funções de ranqueamento quanto à capacidade de diferenciar ativos de inativos pode ser feita usando as mesmas métricas já descritas para avaliar o desempenho dos modelos farmacofóricos.¹²⁴

Um exemplo recente de sucesso do uso do ancoramento foi a busca de novos ativadores do receptor de opioide- μ (μ OR) partindo de um banco de 3 milhões de compostos líder-similar (*lead-like*).¹⁵⁵ Estes compostos foram ancorados no sítio de ligação do receptor μ OR usando o programa *DOCK*.¹⁵⁶ As 2.500 moléculas com melhores valores de escore foram analisadas visualmente, considerando sua novidade estrutural, sua capacidade de formar interações com resíduos polares reconhecidos como importantes para formação do complexo ligante-receptor e excluindo aquelas com conformações energeticamente desfavoráveis. Nesta etapa, 23 moléculas foram selecionadas para serem compradas e testadas frente ao receptor. Destas moléculas, 7 (30%) mostraram K_i entre 2,3 e 14 $\mu\text{mol L}^{-1}$, sendo estruturalmente diferentes dos compostos já descritos como inibidores do receptor μ OR. Entretanto, os compostos inicialmente identificados precisavam ser melhorados quanto às suas atividades (K_i na faixa de poucos $\mu\text{mol L}^{-1}$). Para isso, os autores usaram um conjunto de 500 compostos análogos aos mais ativos contendo as principais regiões para ligação com o receptor, bem como outros grupos capazes de fazer novas interações com o receptor. Estes 500 análogos foram ancorados e os 15 com melhores escores foram selecionados e testados. Destes, 7 (47%) foram ativos, com K_i entre 42 nmol L^{-1} e 4,7 $\mu\text{mol L}^{-1}$. O composto mais ativo (42 nmol L^{-1}) foi,

então, modificado por síntese orgânica, resultando em um composto bastante ativo, com $K_i = 1 \text{ nmol L}^{-1}$, que também apresentou uma alta seletividade. Consideramos importante ressaltar que neste exemplo a seleção final dos compostos foi feita com base na inspeção visual, que é uma etapa indispensável do ancoramento.

Além da inspeção visual, outros aspectos importantes para o sucesso do ancoramento são: (i) a escolha da estrutura tridimensional do alvo molecular a ser usada (alta qualidade), (ii) a inclusão da flexibilidade conformacional do alvo e (iii) a identificação de águas estruturais. Estes aspectos serão brevemente discutidos a seguir:

- (i) Escolha correta da estrutura tridimensional do alvo molecular para busca virtual. É uma etapa crucial em qualquer abordagem baseada na estrutura do receptor, independentemente de como esta estrutura foi obtida (cristalografia de raios-X, RMN e/ou modelagem por homologia). Em qualquer um destes casos, deve-se lembrar que o conjunto de coordenadas usado sempre será um modelo.¹⁵⁷ Assim, a qualidade destes modelos precisa ser verificada.^{151,158} Considerando, por exemplo, as estruturas obtidas por cristalografia de raios-X, alguns dos critérios que deveriam ser observados são: a resolução (medida teórica da menor distância observada no cristal, indicando a quantidade de dados coletados); o *R-factor* (medida da diferença entre os dados observados e os preditos usando o modelo, indicador da qualidade do modelo); *Rfree* (coeficiente de validação cruzada do *R-factor* calculado a partir de dados não usados para construir o modelo); diferença entre *R-factor* e *Rfree* (medida do *overfit* do modelo); contatos cristalinos com o ligante e/ou com o sítio de ligação; e ainda a presença de regiões com densidade incompleta e/ou modelada incorretamente próxima ao ligante e/ou ao sítio de ligação. Como pode ser observado, na análise da qualidade de uma estrutura 3D vários critérios deveriam ser analisados. No entanto, não é raro encontrar na literatura estudos *SBDD* usando estruturas de baixa qualidade.^{64,151,157}
- (ii) Inclusão da flexibilidade conformacional. Até o momento, não é possível incluir todos os graus de liberdade conformacionais de um alvo molecular. Algumas alternativas para contornar esta limitação são: a) atribuir flexibilidade para a cadeia lateral de alguns resíduos de aminoácidos, usando uma biblioteca de rotâmeros e/ou b) usar um conjunto de diferentes conformações do receptor, obtidas tanto experimentalmente quanto por meio de simulações de DM.^{48,50,140}
- (iii) Identificação de águas estruturais. Estas são moléculas de água que fazem ligações de hidrogênio com resíduos da superfície do receptor em uma geometria perfeita, podendo intermediar interações ligante-receptor.²⁹ Em qualquer interação ligante-receptor ocorre o deslocamento de moléculas de água nas superfícies polares do receptor. A dessolvatação de regiões polares tem um custo entálpico que deve ser compensado pela formação de ligações de hidrogênio no complexo ligante-receptor. No entanto, a perda entálpica associada ao deslocamento de moléculas de água estruturais dificilmente é compensada.^{159,160} Uma prática comum nas abordagens *SBDD*, incluindo o ancoramento, é remover todas as moléculas de água presentes na estrutura. No entanto, moléculas de água estruturais deveriam ser mantidas. Exemplos da importância de se considerarem moléculas de água, possivelmente estruturais, no ancoramento já foram descritos na literatura.¹⁶¹ Nos últimos anos, apesar dos esforços feitos para desenvolver modelos capazes de identificar quais moléculas de água são ou não estruturais (ex. *WaterMap*¹⁶²), a identificação e a inclusão destas moléculas de água em estudos de *SBDD* é ainda um grande desafio.^{161,163–165}

Os três aspectos descritos acima são importantes não só nos estudos de ancoramento, mas em qualquer abordagem que considere a estrutura do receptor. Embora os programas de ancoramento tenham

várias limitações, seu uso como filtro na busca virtual resultou na identificação de um grande número de ligantes que foram, subseqüentemente, validados tanto usando ensaios enzimáticos¹⁵⁵ quanto determinando experimentalmente a estrutura 3D do complexo receptor-ligante.^{46,155} Ainda, o sucesso do uso deste filtro vai depender da qualidade do banco de moléculas virtuais, do tipo de alvo e dos critérios de escolha dos *hits*. Aspectos relacionados à qualidade do banco de dados serão apresentados e discutidos a seguir. Em publicação recente, Irwin e Shoichet⁴⁶ sugerem que se poderiam encontrar entre 2 e 5 novas classes de compostos com atividade biológica usando o ancoramento e quando são testados pelo menos de 25 a 50 compostos.⁴⁶ No entanto, a aplicação deste ancoramento a alvos moleculares com sítios de ligação amplos e rasos leva a uma menor taxa de sucesso como, por exemplo, para algumas proteases e interações proteína-proteína.⁴⁶ Nestes casos, seria necessário testar um número muito maior de compostos sem, ainda, garantia de sucesso.⁴⁶

Por fim, independentemente do alvo molecular estudado, para o sucesso do ancoramento é necessário que o pesquisador seja bastante criterioso ao analisar as poses e, principalmente, os valores de escore. Neste contexto, a etapa de inspeção visual é indispensável.⁴⁶ Ainda, o filtro de ancoramento tem o maior custo computacional quando comparado aos filtros apresentados anteriormente (similaridade 2D, farmacofórico e baseado no formato). Assim, na busca virtual usando diferentes filtros de seleção de modo sequencial, o filtro de ancoramento deveria ser utilizado por último, quando o número de compostos já foi significativamente reduzido.¹²⁷

BANCOS DE DADOS VIRTUAIS DE MOLÉCULAS

A preparação e/ou escolha do banco de dados de moléculas a ser usado na busca virtual são etapas cruciais para o sucesso desta busca. Existem vários bancos de dados virtuais de moléculas, gratuitos ou comerciais, que podem ser usados na busca virtual, como o ChEMBL,¹⁶⁶ o ZINC,¹⁶⁷ o PubChem,¹⁶⁸ o PDBbind,¹⁶⁹ o WOMBAT,¹⁷⁰ e o DrugBank,¹⁷¹ dentre outros. O formato das moléculas pode variar de acordo com o banco, por exemplo, uni- (*SMILE*) ou tridimensional (*MOL2* ou *SDF*).

A seqüência de etapas envolvidas na preparação do banco de dados depende de diversos fatores, como o alvo molecular em estudo e o(s) tipo(s) de filtro(s) que será(ão) usado(s) na busca virtual (ex. de similaridade 2D, farmacofórico, baseado no formato e de ancoramento). De modo geral, as principais etapas empregadas na preparação do banco de dados envolvem: remoção de compostos com grupos funcionais indesejados (“*garbage filters*”), remoção de compostos com propriedades físico-químicas e/ou estruturais indesejadas, correção das estruturas e obtenção de diferentes conformações. Cada uma destas etapas será brevemente discutida a seguir.

Remoção de compostos com grupos funcionais indesejados

Os bancos de dados usados para busca virtual podem conter, por exemplo, compostos com grupos funcionais e/ou subestruturas capazes de gerar resultados falso-positivos nos ensaios *in vitro* (ex. como os *PAINS*) e/ou compostos com grupos reativos frente a determinados alvos biológicos. Esta etapa de remoção de compostos, geralmente, é denominada de “*limpeza*” (*cleaning up*) e os filtros usados podem ser chamados de “*garbage filters*”.^{37,64,172} No entanto, como já apresentado e discutido, a aplicação indiscriminada destes filtros, em especial daqueles descritos como *PAINS*,⁹⁰ não é recomendada.^{102,103} Já o uso dos filtros considerando apenas grupos reativos é menos criticado,^{37,64,172} exceto quando o objetivo da busca virtual for selecionar compostos que atuem formando, por exemplo, uma ligação covalente.¹²⁸

Remoção de compostos com propriedades físico-químicas e/ou estruturais indesejadas

Como já apresentado, não há um consenso na literatura sobre a inclusão de filtros de propriedades “fármaco-similar” durante a etapa de preparação dos bancos de dados. Um argumento a favor de incluir estes filtros é poder selecionar, já nas primeiras etapas, compostos com uma maior probabilidade de apresentar propriedades ADMET nas faixas adequadas.^{36,37,97,172} A principal crítica é que a inclusão destes filtros restringe o espaço químico, diminuindo a chance de se encontrarem novas classes de compostos.^{87,173} Outra crítica é que os critérios de corte usados nestes filtros não se aplicam a determinadas classes de compostos (em especial, produtos naturais), bem como a certos tipos de alvos moleculares.^{87,88}

Correção das estruturas

Esta etapa é fundamental durante a preparação dos bancos de dados e não há críticas a respeito da sua inclusão. A etapa de correção de estruturas visa a garantir que as moléculas presentes no banco de dados estejam descritas corretamente quanto à sua quiralidade (quando conhecida); hibridização dos átomos nas ligações químicas; seu estado de protonação (definido para um determinado pH) e quanto à sua forma tautomérica. A relevância desta etapa é facilmente compreendida considerando, principalmente, os filtros baseados nas interações ligante-receptor (como, por exemplo, farmacofórico e ancoragem). Nestes filtros, os padrões de interação de uma determinada molécula poderão depender, por exemplo, de seu estado de protonação (ex. nas interações iônicas) e/ou da sua forma tautomérica (ex. nas ligações de hidrogênio). Além disso, diferentes isômeros estéricos apresentam uma distribuição espacial diferente, que poderá modificar tanto sua forma quanto seu padrão de interação em três dimensões.^{37,64,174}

Obtenção de diferentes conformações

Um dos maiores desafios desta etapa é gerar para cada molécula um conjunto de diferentes conformações que seja adequado para varrer e que represente todo seu espaço conformacional.^{37,64,175} Esta etapa tenta garantir que seja possível encontrar a conformação bioativa dentre as conformações geradas. Existem vários programas capazes de gerar conformações 3D, entre eles: o *Open Babel*,¹⁰⁷ o *Omega*¹⁷⁶ e o *iCon*.¹⁷⁷ Recentemente, encontra-se na literatura uma revisão sobre os métodos usados para gerar estas conformações.¹⁷⁸ No entanto, nenhum deles é capaz de gerar, em todos os casos, conformações próximas às bioativas.¹⁷⁹ O número de conformações necessárias para se varrer o espaço conformacional vai depender da flexibilidade de cada molécula (expressa, geralmente, pelo número de ligações com livre rotação), sendo maior para moléculas mais flexíveis. Assim, num primeiro momento, poderíamos supor que o ideal seria gerar o maior número de possíveis conformações. Entretanto, quanto maior o número de conformações geradas, maior o espaço em disco necessário para armazenar o banco de dados, bem como o custo computacional de aplicar os filtros de seleção.⁶⁴ Não existe, dentro do nosso conhecimento, um número ideal de conformações que devam ser geradas. Além do número de conformações, outro fator importante é evitar a presença daqueles com alta energia. Embora existam na literatura exemplos de ligantes em complexos que apresentam conformações de alta energia, estes não são a maioria, sendo reportado na literatura que 70% dos ligantes formam complexos com energias torcionais inferiores a 3 kcal mol⁻¹.⁶⁴

Na preparação do banco de dados, além das etapas descritas acima, alguns pesquisadores sugerem agrupar as moléculas segundo

suas similaridades estruturais e selecionar algumas como representativas de cada grupo.³⁷

SIMULAÇÕES DE DINÂMICA MOLECULAR APLICADAS NA BUSCA VIRTUAL

Um dos primeiros modelos propostos para explicar a atividade biológica de um determinado composto foi o modelo chave-fechadura descrito por E. Fisher em 1894. Neste modelo, o receptor é a fechadura e o ligante é a chave.¹⁸⁰ As ranhuras da chave corresponderiam às interações ligante-receptor e ambos são considerados como estruturas rígidas.¹⁸⁰ Em 1958, Koshland¹⁸¹ propôs o modelo do ajuste induzido, no qual a formação do complexo ligante-receptor induz mudanças conformacionais no receptor, incluindo assim a sua flexibilidade. Posteriormente, surgiu o modelo de seleção conformacional,¹⁸² no qual o receptor livre (sem o ligante) adota um conjunto de diferentes conformações e o ligante se ligaria, preferencialmente, a uma das conformações pré-existentes. Este modelo foi proposto para explicar as mudanças conformacionais observadas no sítio de ligação após um determinado ligante interagir com um sítio alostérico.^{49,183} Independentemente do modelo considerar que a interação ligação-receptor ocorra por ajuste induzido e/ou seleção conformacional, a flexibilidade do ligante (“chave”) e do receptor (“fechadura”) tornam-se evidentes.¹⁸⁴

De um modo ou de outro, a flexibilidade do ligante já tem sido considerada há algum tempo na busca virtual.^{46,64,140} No entanto, considerar a flexibilidade do receptor continua sendo um desafio.^{46,50,51,184} Como já apresentado, uma estratégia bastante aceita para se incluir a flexibilidade do receptor na busca virtual é usar um conjunto de conformações deste receptor.^{29,48} Este conjunto de conformações pode ser obtido experimental ou computacionalmente, usando Mecânica Estatística (simulações de DM ou MC). Considerando que a obtenção experimental deste conjunto de conformações nem sempre é possível, as simulações usando Mecânica Estatística ganharam uma grande importância, em especial as simulações de DM.^{48,184}

Simulações de DM clássicas são métodos computacionais usados para estudar o movimento dos átomos e das moléculas aplicando a equação de movimento de Newton (Equação 3). Nesta equação, a força líquida (f_i) atuando em um átomo i de um sistema num determinado instante corresponde à sua massa (m) multiplicada por sua aceleração (a). A configuração do sistema neste instante é representada por um vetor r , que descreve a posição de N átomos que interagem no espaço cartesiano ($r = \{x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N\}$). Nas simulações de DM, esta força é descrita de forma clássica usando uma função de energia potencial empírica ($V(r)$, também chamada de campo de força, mostrado na Figura 3S). Nesta aproximação, o movimento dos núcleos é mantido e valores médios são usados para descrever o movimento dos elétrons.

$$\vec{f}_i = m_i \vec{a}_i = - \frac{\partial V(\vec{r})}{\partial \vec{r}_i} \quad (3)$$

Assim, a integração da equação de movimento de Newton durante a simulação gera sucessivas configurações do sistema, resultando numa trajetória de coordenadas em função do tempo.⁴⁸ Uma limitação nestes campos de força é não considerar a polarizabilidade dos átomos, pois as cargas parciais destes não podem se alterar durante a simulação. Campos de força polarizáveis foram criados para contornar esta limitação, mas são ainda muito caros computacionalmente. Outra limitação desses campos de força é não permitir que ocorra a quebra ou formação de novas ligações, impedindo que mecanismos de catálise enzimática sejam estudados.¹⁸⁵ Para contornar esta limitação podem-se usar simulações híbridas com mecânica quântica e clássica (QM/MM).

As trajetórias obtidas nas simulações de DM são um conjunto *snapshots* (fotografias). Os *snapshots* representam diferentes conformações do receptor ao longo do tempo. Estas conformações podem ser agrupadas selecionando-se algumas conformações representativas, usando, por exemplo, os valores de *RMSD* da cadeia principal do receptor calculados ao longo de toda trajetória. Estas conformações podem, então, ser usadas na busca virtual, aplicando-se, por exemplo, filtro de ancoramento.^{48,51,186} No caso de simulações da estrutura do receptor complexada com o ligante, os diferentes modos de ligações do ligante observados durante a trajetória podem ser usados para gerar diferentes modelos farmacofóricos.^{51,132,133} Além disso, a estabilidade da(s) interação(ões) ligante-receptor pode ser analisada ao longo da trajetória e utilizada para auxiliar na escolha das regiões de interação presentes nos modelos farmacofóricos.¹³²

CONCLUSÕES

A busca virtual de compostos bioativos tem sido bastante utilizada na identificação de novos compostos com atividade biológica, em especial de fármacos. Esta busca consiste em pré-selecionar compostos com o auxílio do computador a partir de bancos de dados virtuais com um grande número de moléculas. No entanto, a proposição de modelos de busca virtual com alta taxa de acerto é uma tarefa ainda bastante desafiadora, uma vez que este modelo considera um sistema inerentemente muito complexo. Assim, uma das estratégias sugeridas na literatura para realizar a busca virtual é a utilização sequencial de diferentes filtros de seleção, sendo que o sucesso desta busca está fortemente relacionado com a qualidade e eficiência de cada um dos filtros aplicados, bem como do banco de dados usado.

Neste artigo de revisão descrevemos alguns dos filtros de seleção que podem ser usados na busca virtual (filtros de propriedades fármaco-similar, de características estruturais, de similaridade 2D, farmacofórico, baseado no formato e de ancoramento). Para cada um destes filtros descrevemos suas características, vantagens e limitações, acompanhadas de literatura recente e relevante para a Química Medicinal, área que tem mostrado significativo avanço no Brasil.¹⁸⁷ Adicionalmente, apresentamos as principais etapas envolvidas na preparação de um banco de dados virtual de moléculas, acompanhadas de literatura recente. Esperamos que este artigo de revisão possa contribuir para aprimorar ou complementar os conhecimentos do leitor, em aspectos que consideramos relevantes para compreender cada uma das estratégias usadas em Química Medicinal aqui apresentadas.

MATERIAL SUPLEMENTAR

Algumas imagens das estruturas/grupos funcionais usados como filtros estruturais e da representação do campo de força utilizados neste trabalho estão disponíveis em <http://quimicanova.sbq.org.br>, na forma de arquivo PDF, com acesso livre.

AGRADECIMENTOS

Agradecemos aos Professores E. I. Ferreira e R. V. C. Guido por nos alertar da importância de disponibilizar este material em português para a comunidade científica, em especial para os alunos iniciantes na área. E, ainda, à OpenEye Scientific Software e à Inte:Ligand pelos programas computacionais usados para gerar as Figuras 3, 5 e 6. Este trabalho foi financiado pela Fundação de Amparo à Pesquisa no Estado de São Paulo (FAPESP) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A.T-do Amaral é membro do CEPID Redoxoma (No. 2013/07937-8) e do NAP Redoxoma (PRPUSP). E. Piccirillo foi bolsista de Doutorado da FAPESP (No. 2014/01614-5 e 2012/06633-2).

REFERÊNCIAS E NOTAS

- Ravina, E.; *The Evolution of Drug Discovery: From Traditional Medicines to Modern Drugs* John Wiley & Sons: New York, 2011.
- Viegas, C.; Bolzani, V. da S.; Barreiro, E. J.; *Quim. Nova* **2006**, *29*, 326.
- Malvezzi, A.; *Tese de Doutorado*, Universidade de São Paulo, São Paulo, **2008**.
- Kubinyi, H.; *Industrial Pharmacy* **2004**, *7*.
- Faller, B.; Ottaviani, G.; Ertl, P.; Berellini, G.; Collis, A.; *Drug Discovery Today* **2011**, *16*, 976.
- Kubinyi, H.; *J. Recept. Signal Transduct. Res.* **1999**, *19*, 15.
- Kubinyi, H.; *Nat. Rev. Drug Discovery* **2003**, *2*, 665.
- Rollinger, J. M.; Stuppner, H.; Langer, T.; *Prog. Drug Res.* **2008**, *65*, 211.
- Cumming, J. G.; Davis, A. M.; Muresan, S.; Haeblerlein, M.; Chen, H.; *Nat. Rev. Drug Discovery* **2013**, *12*, 948.
- Mignani, S.; Huber, S.; Tomás, H.; Rodrigues, J.; Majoral, J. P.; *Drug Discovery Today* **2016**, *21*, 573.
- Kubinyi, H.; Em *Computer Applications in Pharmaceutical Research and Development*; Ekins, S. ed.; 1ed., John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006, cap. 16.
- Kubinyi, H.; *Proceedings of the Beilstein Bozen symposium – The Chemical Theatre of Biological Systems*, Bozen, Italy, **2004**, 51.
- Kubinyi, H.; Em *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*; Stroud, R. M.; Finer-Moore, J. eds.; 1ed., Royal Society of Chemistry: Cambridge, UK, 2007, cap. 2.
- Hansch, C.; Fujita, T.; *J. Am. Chem. Soc.* **1964**, *86*, 1616.
- Kubinyi, H.; *QSAR: Hansch Analysis and Related Approaches* Wiley-VCH Verlag GmbH: Weinheim, **1993**.
- Free, S. M.; Wilson, J. W.; *J. Med. Chem.* **1964**, *7*, 395.
- Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; *Anal. Chim. Acta* **2004**, *515*, 199.
- Topliss, J. G.; Em *Computer Aided Drug Design in Industrial Research*; Herrmann, E. C.; Franke, R. eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **1995**, cap. 2.
- Fujita, T.; Winkler, D. A.; *J. Chem. Inf. Model.* **2016**, *56*, 269.
- Cramer, R. D.; *J. Comput. Aided. Mol. Des.* **2012**, *26*, 35.
- Gola, J.; Obrezanova, O.; Champness, E.; Segall, M.; *QSAR Comb. Sci.* **2006**, *25*, 1172.
- Luo, M.; Wang, X. S.; Roth, B. L.; Golbraikh, A.; Tropsha, A.; *J. Chem. Inf. Model.* **2014**, *54*, 634.
- 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.; Terflath, L.; Gasteiger, J.; Richard, A.; Tropsha, A.; *J. Med. Chem.* **2014**, *57*, 4977.
- Kubinyi, H.; *Drug Discovery Today* **1997**, *2*, 538.
- Dias, R. L. A.; Corrêa, A. G.; *Quim. Nova* **2001**, *24*, 236.
- Ferreira, R. S.; Glaucius, O.; Andricopulo, A. D.; *Quim. Nova* **2011**, *34*, 1770.
- Mayr, L. M.; Bojanic, D.; *Curr. Opin. Pharmacol.* **2009**, *9*, 580.
- Kirchmair, J.; Markt, P.; Distinto, S.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Langer, T.; Wolber, G.; *J. Med. Chem.* **2008**, *51*, 7021.
- Klebe, G.; *Drug Discovery Today* **2006**, *11*, 580.
- Chen, W. N.; Nitsche, C.; Pilla, K. B.; Graham, B.; Huber, T.; Klein, C. D.; Otting, G.; *J. Am. Chem. Soc.* **2016**, *138*, 4539.
- Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C.; *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899.
- Behnen, J.; Köster, H.; Neudert, G.; Craan, T.; Heine, A.; Klebe, G.; *ChemMedChem* **2012**, *7*, 248.

33. Glen, R. C.; *Future Med. Chem.* **2011**, *3*, 399.
34. Hopkins, A. L.; Groom, C. R.; *Nat. Rev. Drug Discovery* **2002**, *1*, 727.
35. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.; *Adv. Drug Deliv. Rev.* **2001**, *46*, 3.
36. Meanwell, N. A.; *Chem. Res. Toxicol.* **2011**, *24*, 1420.
37. Bologa, C. G.; Oprea, T. I.; Em *Bioinformatics and Drug Discovery*; Larson, R. S. ed.; Humana Press: Totowa, 2012, cap. 7.
38. Kirchmair, J.; Göller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G.; *Nat. Rev. Drug Discovery* **2015**, *14*, 387.
39. Rondeau, J.-M.; Schreuder, H.; Em *The Practice of Medicinal Chemistry; Wermuth, C. G. ed.*; 3^a ed.; Elsevier: Amsterdam, 2008, cap. 30.
40. Kawakami, Y.; Inoue, A.; Kawai, T.; Wakita, M.; Sugimoto, H.; Hopfinger, A. J.; *Bioorg. Med. Chem.* **1996**, *4*, 1429.
41. von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R.; *Nature* **1993**, *363*, 418.
42. Wade, R. C.; *Structure* **1997**, *5*, 1139.
43. Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A.; *Nat. Rev. Drug Discovery* **2002**, *1*, 493.
44. Maibaum, J.; Feldman, D. L.; *Annu. Rep. Med. Chem.* **2009**, *44*, 105.
45. Weber, A. E.; Thornberry, N.; *Annu. Rep. Med. Chem.* **2007**, *42*, 95.
46. Irwin, J. J.; Shoichet, B. K.; *J. Med. Chem.* **2016**, *59*, 4103.
47. Swinney, D. C.; Anthony, J.; *Nat. Rev. Drug Discovery* **2011**, *10*, 507.
48. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A.; *J. Med. Chem.* **2016**, *59*, 4035.
49. Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C.; *Acc. Chem. Res.* **2016**, *49*, 809.
50. Sottriffer, C. A.; *Curr. Top. Med. Chem.* **2011**, *11*, 179.
51. Wieder, M.; Garon, A.; Perricone, U.; Boresch, S.; Seidel, T.; Almerico, A. M.; Langer, T.; *J. Chem. Inf. Model.* **2017**, *57*, 365.
52. *Hit* é um composto com estrutura identificada/conhecida que possui uma atividade em um ensaio biológico abaixo de determinado valor corte.
53. *Lead* é um composto (ou uma série de compostos) com atividade e seletividade comprovada através de ensaios biológicos que, ainda, preenchem alguns critérios importantes para desenvolvimento de um fármaco como originalidade, patentabilidade e acessibilidade.
54. Brown, F. K.; Sherer, E. C.; Johnson, S. A.; Holloway, M. K.; Sherborne, B. S.; *J. Comput. Aided. Mol. Des.* **2017**, *31*, 255.
55. Muegge, I.; Bergner, A.; Kriegl, J. M.; *J. Comput. Aided. Mol. Des.* **2017**, *31*, 275.
56. McGaughey, G.; Patrick Walters, W.; *J. Comput. Aided. Mol. Des.* **2017**, *31*, 293.
57. Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Körner, M.; Kuglstatler, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M.; *J. Med. Chem.* **2016**, *59*, 4087.
58. Gohlke, H.; Hendlich, M.; Klebe, G.; *J. Mol. Biol.* **2000**, *295*, 337.
59. Lipinski, C. A.; *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235.
60. Wenlock, M. C.; Barton, P.; *Mol. Pharm.* **2013**, *10*, 1224.
61. Leach, A. R.; Hann, M. M.; *Curr. Opin. Chem. Biol.* **2011**, *15*, 489.
62. Piccirillo, E.; Amaral, A. T.; *Quim. Nova*, dados não publicados.
63. Mattos, C.; Ringe, D.; Em *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H. ed.; Springer: Berlin, **1993**.
64. Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K.; *J. Chem. Inf. Model.* **2012**, *52*, 867.
65. Koch, O.; Cappel, D.; Nocker, M.; Jäger, T.; Flohé, L.; Sottriffer, C. A.; Selzer, P. M.; *PLoS One* **2013**, *8*, e56788.
66. Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J.; *Nat. Rev. Drug Discovery* **2011**, *10*, 197.
67. Hodgson, J.; *Nat. Biotechnol.* **2001**, *19*, 722.
68. Oprea, T. I.; *Molecules* **2002**, *7*, 51.
69. Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K.; *J. Med. Chem.* **2010**, *53*, 4891.
70. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.; *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
71. Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T.; *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 3743.
72. Makara, G. M.; *J. Med. Chem.* **2007**, *50*, 3214.
73. Goldberg, F. W.; Kettle, J. G.; Kogej, T.; Perry, M. W. D.; Tomkinson, N. P.; *Drug Discovery Today* **2015**, *20*, 11.
74. Tice, C. M.; *Pest Manag. Sci.* **2001**, *57*, 3.
75. Ghose, A. K.; Crippen, G. M.; *J. Comput. Chem.* **1986**, *7*, 565.
76. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D.; *J. Med. Chem.* **2002**, *45*, 2615.
77. Martin, Y. C.; *J. Med. Chem.* **2005**, *48*, 3164.
78. DeGoey, D. A.; Chen, H.-J.; Cox, P. B.; Wendt, M. D.; *J. Med. Chem.* **2017**, doi: 10.1021/acs.jmedchem.7b00717.
79. Tóth, R.; van der Hoor, R. A. L.; *Trends Plant Sci.* **2010**, *15*, 81.
80. Hill, A. P.; Young, R. J.; *Drug Discovery Today* **2010**, *15*, 648.
81. Mannhold, R.; Berellini, G.; Carosati, E.; Benedetti, P.; Em *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*; Cruciani, G. ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2006, cap. 8.
82. Chevillard, F.; Lagorce, D.; Reynès, C.; Villoutreix, B. O.; Vayer, P.; Miteva, M. A.; *Mol. Pharm.* **2012**, *9*, 3127.
83. Cruciani, G.; Crivori, P.; Carrupt, P. A.; Testa, B.; *J. Mol. Struct. (Theochem)* **2000**, *503*, 17.
84. Kenny, P. W.; Montanari, C. A.; *J. Comput. Aided. Mol. Des.* **2013**, *27*, 1.
85. Sant'Anna, C. M. R.; *Quim. Nova* **2002**, *25*, 505.
86. Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J.; *Nat. Rev. Drug Discovery* **2015**, *14*, 111.
87. Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J.; *J. Med. Chem.* **2016**, *59*, 2312.
88. Lipinski, C. A.; *Adv. Drug Deliv. Rev.* **2016**, *101*, 34.
89. Baell, J.; Walters, M. A.; *Nature* **2014**, *513*, 481.
90. Baell, J. B.; Holloway, G. A.; *J. Med. Chem.* **2010**, *53*, 2719.
91. Rishton, G. M.; *Drug Discovery Today* **2003**, *8*, 86.
92. Jadhav, A.; Ferreira, R. S.; Klumpp, C.; Mott, B. T.; Austin, C. P.; Inglese, J.; Thomas, C. J.; Maloney, D. J.; Shoichet, B. K.; Simeonov, A.; *J. Med. Chem.* **2010**, *53*, 37.
93. Owens, P. K.; Raddad, E.; Miller, J. W.; Stille, J. R.; Olovich, K. G.; Smith, N. V.; Jones, R. S.; Scherer, J. C.; *Nat. Rev. Drug Discovery* **2014**, *14*, 17.
94. Malvezzi, A.; de Rezende, L.; Izidoro, M. A.; Cezari, M. H. S.; Juliano, L.; Amaral, A. T.-do.; *Bioorg. Med. Chem. Lett.* **2008**, *18*, 350.
95. Clark, D. E.; Pickett, S. D.; *Drug Discovery Today* **2000**, *5*, 49.
96. Mullard, A.; *Nat. Rev. Drug Discovery* **2014**, *13*, 877.
97. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A.; *Nat. Rev. Drug Discovery* **2015**, *14*, 475.
98. Schulze, U.; Baedeker, M.; Chen, Y. T.; Greber, D.; *Nat. Rev. Drug Discovery* **2014**, *13*, 331.
99. Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K.; *J. Med. Chem.* **2015**, *58*, 7076.
100. Rishton, G. M.; *Drug Discovery Today* **1997**, *2*, 382.
101. Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjögren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G.; *J. Med. Chem.* **2002**, *45*, 137.
102. Capuzzi, S. J.; Muratov, E. N.; Tropsha, A.; *J. Chem. Inf. Model.* **2017**, *57*, 417.
103. Jasial, S.; Hu, Y.; Bajorath, J.; *J. Med. Chem.* **2017**, *60*, 3879.

104. Piccirillo, E.; *Tese de Doutorado*, Universidade de São Paulo, Brasil, 2017.
105. Kubinyi, H.; *Perspect. Drug Discovery Des.* **1998**, *9*, 225.
106. Willett, P.; *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2011**, *1*, 241.
107. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R.; *J. Cheminform.* **2011**, *3*, 33.
108. Rogers, D.; Hahn, M.; *J. Chem. Inf. Model.* **2010**, *50*, 742.
109. O'Boyle, N. M.; Sayle, R. A.; *J. Cheminform.* **2016**, *8*, 1.
110. Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G.; *Methods* **2015**, *71*, 58.
111. Stumpfe, D.; Bajorath, J.; *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 260.
112. Willett, P.; *Drug Discovery Today* **2006**, *11*, 1046.
113. Stumpfe, D.; Hu, Y.; Dimova, D.; Bajorath, J.; *J. Med. Chem.* **2014**, *57*, 18.
114. Maggiora, G. M.; *J. Chem. Inf. Model.* **2006**, *46*, 1535.
115. Bajorath, J.; *Expert Opin. Drug Discovery* **2017**, *12*, 879.
116. Sippl, W.; Em *The Practice of Medicinal Chemistry*; Wermuth, C. G., ed.; 3^a ed., Elsevier Ltd: Amsterdam, 2008, cap. 8.
117. Wolber, G.; Langer, T.; *J. Chem. Inf. Model.* **2005**, *45*, 160.
118. Natesh, R.; Schwager, S. L. U.; Sturrock, E. D.; Acharya, K. R.; *Nature* **2003**, *421*, 1427.
119. Markt, P.; Schuster, D.; Langer, T.; Em *Virtual Screening - Principles, Challenges, and Practical Guidelines*; Sotriffer, C. A., ed.; 1^a ed., Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2011, cap. 5.
120. Kuhnert, M.; Diederich, W. E.; *Synlett* **2016**, *27*, 641.
121. Catalyst; *Pharmacophore Modeling and Analysis tools*, Accelrys Software Inc., San Diego, CA, USA, **2005**.
122. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S.; *J. Chem. Inf. Model.* **2007**, *47*, 279.
123. ChemicalComputingGroupInc.; *Sci. Comput. Instrum.* **2004**, *22*, 32.
124. McGann, M.; Nicholls, A.; Enyedy, I.; *J. Comput. Aided. Mol. Des.* **2015**, *29*, 923.
125. Hristozov, D. P.; Oprea, T. I.; Gasteiger, J.; *J. Comput. Aided. Mol. Des.* **2007**, *21*, 617.
126. Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T.; *J. Comput. Aided. Mol. Des.* **2008**, *22*, 213.
127. Malvezzi, A.; Queiroz, R. F.; de Rezende, L.; Augusto, O.; do Amaral, A. T.; *Mol. Inform.* **2011**, *30*, 605.
128. Malvezzi, A.; de Rezende, L.; do Amaral, A. T.; *QSAR Comb. Sci.* **2009**, *28*, 781.
129. Rechfeld, F.; Gruber, P.; Kirchmair, J.; Boehler, M.; Hauser, N.; Hechenberger, G.; Garczarczyk, D.; Lapa, G. B.; Preobrazhenskaya, M. N.; Goekjian, P.; Langer, T.; Hofmann, J.; *J. Med. Chem.* **2014**, *57*, 3235.
130. Hu, B.; Lill, M. A.; *J. Chem. Inf. Model.* **2013**, *53*, 1179.
131. Goodford, P. J.; *J. Med. Chem.* **1985**, *28*, 849.
132. Piccirillo, E.; Merget, B.; Sotriffer, C. A.; Do Amaral, A. T.; *J. Comput. Aided. Mol. Des.* **2016**, *30*, 251.
133. Wieder, M.; Perricone, U.; Seidel, T.; Boesch, S.; Langer, T.; *Monatsh. Chem.* **2016**, *147*, 553.
134. Choudhury, C.; Priyakumar, U. D.; Sastry, G. N.; *J. Chem. Inf. Model.* **2015**, *55*, 848.
135. Nicholls, A.; McGaughey, G. B.; Sheridan, R. P.; Good, A. C.; Warren, G.; Mathieu, M.; Muchmore, S. W.; Brown, S. P.; Grant, J. A.; Haigh, J. A.; Nevins, N.; Jain, A. N.; Kelley, B.; *J. Med. Chem.* **2010**, *53*, 3862.
136. Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A.; *J. Med. Chem.* **2007**, *50*, 74.
137. Sastry, G. M.; Dixon, S. L.; Sherman, W.; *J. Chem. Inf. Model.* **2011**, *51*, 2455.
138. Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G.; *J. Chem. Inf. Model.* **2009**, *49*, 678.
139. OpenEye Scientific Software; Santa Fe, N.; <http://www.eyesopen.com>, acessada em Fevereiro 2018.
140. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J.; *Nat. Rev. Drug Discovery* **2004**, *3*, 935.
141. McGann, M.; *J. Chem. Inf. Model.* **2011**, *51*, 578.
142. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L.; *J. Med. Chem.* **2004**, *47*, 1750.
143. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J.; *J. Comput. Chem.* **2009**, *30*, 2785.
144. Trott, O.; Olson, A. J.; *J. Comput. Chem.* **2009**, *31*, 455.
145. Silva, C.; Magalhães, D.; Marinho, D.; José, H.; Barbosa, C.; Emmanuel, L.; *Inf. Sci.* **2014**, *289*, 206.
146. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D.; *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 609.
147. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G.; *J. Mol. Biol.* **1996**, *261*, 470.
148. DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R.; *J. Med. Chem.* **1988**, *31*, 722.
149. RMSD (do inglês *Root Mean Square Distance*) é o valor da raiz quadrada da média dos quadrados das distâncias entre os pares de átomos de duas moléculas.
150. Carlson, H. A.; Smith, R. D.; Damm-Ganamet, K. L.; Stuckey, J. A.; Ahmed, A.; Convery, M. A.; Somers, D. O.; Kranz, M.; Elkins, P. A.; Cui, G.; Peishoff, C. E.; Lambert, M. H.; Dunbar, J. B.; *J. Chem. Inf. Model.* **2016**, *56*, 1063.
151. Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A.; *J. Comput. Aided. Mol. Des.* **2008**, *22*, 179.
152. Korb, O.; Ten Brink, T.; Raj, F. R. D. V. P.; Keil, M.; Exner, T. E.; *J. Comput. Aided. Mol. Des.* **2012**, *26*, 185.
153. Smith, R. D.; Damm-Ganamet, K. L.; Dunbar, J. B.; Ahmed, A.; Chinnaswamy, K.; Delproposito, J. E.; Kubish, G. M.; Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Stuckey, J. A.; Baker, D.; Carlson, H. A.; *J. Chem. Inf. Model.* **2016**, *56*, 1022.
154. Schomburg, K. T.; Rarey, M.; *J. Chem. Inf. Model.* **2014**, *54*, 2261.
155. Manglik, A.; Lin, H.; Aryal, D. K.; McCorvy, J. D.; Dengler, D.; Corder, G.; Levit, A.; Kling, R. C.; Bernat, V.; Hübner, H.; Huang, X.-P.; Sassano, M. F.; Giguère, P. M.; Löber, S.; Da Duan; Scherrer, G.; Kobilka, B. K.; Gmeiner, P.; Roth, B. L.; Shoichet, B. K.; *Nature* **2016**, *537*, 185.
156. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C.; *J. Comput. Chem.* **2015**, *36*, 1132.
157. Warren, G. L.; Do, T. D.; Kelley, B. P.; Nicholls, A.; Warren, S. D.; *Drug Discovery Today* **2012**, *17*, 1270.
158. Dunbar, J. B.; Smith, R. D.; Damm-Ganamet, K. L.; Ahmed, A.; Esposito, E. X.; Delproposito, J.; Chinnaswamy, K.; Kang, Y. N.; Kubish, G.; Gestwicki, J. E.; Stuckey, J. A.; Carlson, H. A.; *J. Chem. Inf. Model.* **2013**, *53*, 1842.
159. Bissantz, C.; Kuhn, B.; Stahl, M.; *J. Med. Chem.* **2010**, *53*, 5061.
160. Ritschel, T.; Kohler, P. C.; Neudert, G.; Heine, A.; Diederich, F.; Klebe, G.; *ChemMedChem* **2009**, *4*, 2012.
161. Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T.; *J. Chem. Theory Comput.* **2014**, *10*, 2769.
162. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A.; *J. Am. Chem. Soc.* **2008**, *130*, 2817.
163. Betz, M.; Wulsdorf, T.; Krimmer, S. G.; Klebe, G.; *J. Chem. Inf. Model.* **2016**, *56*, 223.
164. Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A.; *J. Med. Chem.* **2016**, *59*, 4364.
165. Nittinger, E.; Schneider, N.; Lange, G.; Rarey, M.; *J. Chem. Inf. Model.* **2015**, *55*, 771.
166. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.;

- Papadatos, G.; Santos, R.; Overington, J. P.; *Nucleic Acids Res.* **2014**, *42*, D1083.
167. Sterling, T.; Irwin, J. J.; *J. Chem. Inf. Model.* **2015**, *55*, 2324.
168. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H.; *Nucleic Acids Res.* **2016**, *44*, D1202.
169. Wang, R.; Fang, X.; Lu, Y.; Wang, S.; *J. Med. Chem.* **2004**, *47*, 2977.
170. Olah, M.; Mracec, M.; Ostopovici, L.; Rad, R.; Bora, A.; Hadaruga, N.; Olah, I.; Banda, M.; Simon, Z.; Mracec, M.; Oprea, T. I.; *Em Chemoinformatics in Drug Discovery*; Oprea, T. I.; ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2005, cap. 9.
171. Wishart, D. S.; *Nucleic Acids Res.* **2006**, *34*, D668.
172. Mignani, S.; Huber, S.; Tomás, H.; Rodrigues, J.; Majoral, J. P.; *Drug Discovery Today* **2016**, *21*, 239.
173. Benet, L. Z.; Hosey, C. M.; Ursu, O.; Oprea, T. I.; *Adv. Drug Deliv. Rev.* **2016**, *101*, 89.
174. Martin, Y. C.; *J. Comput. Aided. Mol. Des.* **2009**, *23*, 693.
175. Hawkins, P. C. D.; Nicholls, A.; *J. Chem. Inf. Model.* **2012**, *52*, 2919.
176. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T.; *J. Chem. Inf. Model.* **2010**, *50*, 572.
177. *iCon, Part of LigandScout Version 4.1*; Inte:Ligand: Vienna, Austria, **2017**.
178. Hawkins, P. C. D.; *J. Chem. Inf. Model.* **2017**, *57*, 1747.
179. Friedrich, Nils-Ole; de Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J.; *J. Chem. Inf. Model.* **2017**, doi: 10.1021/acs.jcim.7b00505.
180. Fischer, E.; *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985.
181. Koshland, D. E.; *Proc. Natl. Acad. Sci.* **1958**, *44*, 98.
182. Boehr, D. D.; Nussinov, R.; Wright, P. E.; *Nat. Chem. Biol.* **2009**, *5*, 789.
183. Sítio alostérico é um sítio do receptor distante do seu sítio de ligação, mas ativamente acoplado a ele. Isso significa que a ligação de uma molécula ao sítio alostérico altera o formato do sítio de ligação, modificando a atividade deste receptor.
184. Cozzini, P.; Kellogg, G. E.; Spyrikis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sottriffer, C. A.; *J. Med. Chem.* **2008**, *51*, 6237.
185. Arantes, G. M.; *Quim. Nova* **2008**, *31*, 377.
186. Zhao, H.; Caffisch, A.; *Eur. J. Med. Chem.* **2015**, *91*, 4.
187. do Amaral, A. T.; Andrade, C. H.; Kummerle, A.; Guido, R. V. C.; *Quim. Nova* **2017**, *40*, 694.