

Artigo

Estratégias de Triagem Virtual no Planejamento de Fármacos**Rodrigues, R. P.; Mantoani, S. P.; de Almeida, J. R.; Pinsetta, F. R.; Semighini, E. P.; da Silva, V. B.;*** da Silva, C. H. P.*Rev. Virtual Quim.*, 2012, 4 (6), 739-776. Data de publicação na Web: 24 de dezembro de 2012<http://www.uff.br/rvq>**Virtual Screening Strategies in Drug Design**

Abstract: The development of virtual screening techniques represents a major advance in the current drug design era. Through several strategies, virtual screening is able to facilitate the selection of molecules with the desired chemical features to modulate the biological activity of the most attractive molecular targets currently available. From the simplest techniques, as the similarity search or molecular docking, to more complex strategies, including statistical methods and machine learning, the main goal of virtual screening is to improve the searching for molecules with the desired features required for becoming drug candidates, thus accelerating the continuous process of drug design. The aim of this review is to discuss the main virtual screening strategies and how they relate to the drug design process.

Keywords: Virtual screening; drug design; molecular modeling.

Resumo

O desenvolvimento de técnicas de triagem virtual representa um dos maiores avanços na atual era de planejamento de fármacos. A triagem virtual, através de inúmeras estratégias distintas, é capaz de direcionar a seleção de moléculas com as características químicas desejadas para modular a atividade biológica dos mais diversos e atrativos alvos moleculares conhecidos na atualidade. Desde as técnicas mais simples, como a busca por similaridade ou “docagem” molecular, até as estratégias mais complexas, que envolvem métodos estatísticos e de aprendizagem de máquinas, o objetivo principal da triagem virtual é aprimorar o processo de busca de novos candidatos a fármacos e acelerar o processo contínuo do seu planejamento. O objetivo desta revisão é discutir as principais técnicas de triagem virtual e como elas se relacionam com o desenvolvimento de novos fármacos.

Palavras-chave: Triagem virtual; planejamento de fármacos; modelagem molecular.

* Pontifícia Universidade Católica de Goiás, Departamento de Biomedicina, Rua 232, nº 128, Setor Universitário, CEP: 74605-140, Goiânia, GO, Brasil.

✉ vinciusbarreto11@yahoo.com.br

DOI: [10.5935/1984-6835.20120055](https://doi.org/10.5935/1984-6835.20120055)

Estratégias de Triagem Virtual no Planejamento de Fármacos

Ricardo P. Rodrigues,^a Susimaira P. Mantoani,^a Jonathan R. de Almeida,^a
Flávio Roberto Pinsetta,^a Evandro P. Semighini,^a Vinicius B. da Silva,^{b,*}
Carlos Henrique T. P. da Silva^a

^a Universidade de São Paulo, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do Café, s/n, Monte Alegre, CEP: 14040-903, Ribeirão Preto, SP, Brasil.

^b Pontifícia Universidade Católica de Goiás, Departamento de Biomedicina, Rua 232, nº 128, Setor Universitário, CEP: 74605-140, Goiânia, GO, Brasil.

* viniciusbarreto11@yahoo.com.br

Recebido em 21 de julho de 2012. Aceito para publicação em 12 de dezembro de 2012

1. Introdução
2. Bibliotecas Virtuais
3. Triagem Virtual Baseada na Estrutura do Alvo Molecular
 - 3.1. Algoritmos de busca
 - 3.2. Funções de escore
 - 3.3. Conformação do alvo molecular
4. Triagem Virtual Baseada em Ligantes
 - 4.1. Triagem virtual por padrão farmacofórico
 - 4.2. Triagem por similaridade
 - 4.3. Descritores e novas tendências
 - 4.4. Aprendizagem de máquinas
5. Conclusões

1. Introdução

Nas últimas décadas, um grande número de tecnologias emergentes têm sido aplicadas em projetos de pesquisa e desenvolvimento de fármacos. Neste contexto, destacam-se a genômica e a proteômica, que oferecem informações relevantes de novos e atrativos alvos moleculares e sua relação com os mais

diversos processos fisiopatológicos que afligem a humanidade.¹⁻³ A maior parte dessas tecnologias fornece à comunidade científica grande diversidade de dados e informações que podem ser traduzidas em conhecimento químico e biológico. Com esse propósito, as técnicas computacionais aplicadas ao estudo de sistemas biológicos mostram-se eficazes no manejo de dados e mapeamento da estrutura 3D de alvos moleculares e ligantes, guiando a

identificação e otimização de novos candidatos a fármacos.⁴⁻⁶

Um novo conceito nos projetos de planejamento de fármacos é conhecido como “druggability” e nada mais é do que a possibilidade de modular um alvo biológico lançando mão de moléculas orgânicas, com o objetivo de obter efeitos terapêuticos. Os alvos mais atrativos são os que possuem representatividade na gênese ou progressão de um processo patológico. Com as informações obtidas até o momento através da codificação do genoma humano, o grande desafio é entender as funções de tais alvos moleculares durante os eventos que induzem a progressão das doenças e descobrir substâncias capazes de modular tais eventos e restabelecer a saúde do indivíduo. De acordo com Cheng e colaboradores, cerca de 60% dos projetos de descoberta de novos fármacos baseados em moléculas pequenas falham no processo de transposição da etapa inicial (de promissor ligante) até a etapa final (de candidato a fármaco), pois o alvo biológico em questão não participa diretamente na patologia e, por conseguinte, sua modulação apenas não garantirá a eficácia do tratamento farmacológico⁸.

Desde a concepção do alvo biológico até a descoberta de um novo fármaco, o que pode levar uma década ou mais,⁹ inúmeras abordagens computacionais de bioinformática e quimioinformática podem ser aplicadas nos diferentes estágios do processo de desenvolvimento. Nos estágios iniciais, o foco principal concentra-se em identificar possíveis ligantes para um alvo biológico, enquanto que nos estágios finais a ideia é trabalhar apenas com ligantes que tenham propriedades físico-químicas, farmacodinâmicas e farmacocinéticas ideais para se tornarem fármacos e adentrar o mercado farmacêutico.¹⁰⁻¹²

O grande número de alvos moleculares atrativos explorados atualmente no planejamento de fármacos requer o emprego de métodos rápidos e relativamente precisos capazes de identificar protótipos, também conhecidos como compostos de partida. Uma das principais técnicas utilizadas atualmente

na busca e descoberta de compostos de partida é a triagem virtual.^{6,13-15}

As técnicas de triagem virtual, através da utilização de métodos computacionais, também conhecidos pelo termo “in silico”, auxiliam na seleção de compostos orgânicos promissores como ligantes de alvos terapêuticos de interesse, sejam eles agonistas ou antagonistas de receptores, ou ainda inibidores enzimáticos. Normalmente, a estrutura química de possíveis ligantes advém de bases de dados que contém milhares de compostos com propriedades físico-químicas semelhantes às exibidas por fármacos reconhecidos na prática clínica, além de grande diversidade química e, portanto, preenchendo os requisitos básicos para garantir ampla gama de possíveis interações intermoleculares com os mais diversos alvos moleculares.¹⁶ Dessa forma, os compostos identificados “in silico” como promissores, tanto do ponto de vista farmacofórico como do ponto de vista de suas propriedades físico-químicas, podem ser sintetizados, ou mesmo adquiridos de empresas especializadas e, posteriormente, investigados por ensaios biológicos.

As estratégias de triagem virtual compreendem, basicamente, duas abordagens principais: a triagem baseada na estrutura do alvo molecular e a triagem baseada em ligantes. As técnicas baseadas na estrutura do alvo molecular consideram a estrutura tridimensional (3D) do alvo terapêutico, utilizando como estratégia principal os cálculos de docagem para seleção de ligantes potenciais, com características químicas, eletrônicas e estruturais que favorecem interações com o sítio ligante do alvo molecular. O planejamento de substâncias bioativas com base na estrutura do alvo molecular é uma das estratégias mais robustas de identificação de novos ligantes, capaz de contribuir em todo o processo, desde a análise da estrutura 3D do alvo terapêutico até a otimização das interações moleculares e propriedades farmacocinéticas de compostos candidatos a ensaios clínicos.¹⁷⁻¹⁹

A disponibilidade de informações

estruturais acerca do alvo terapêutico de interesse auxilia na descoberta e planejamento de compostos com complementaridade estérica, eletrostática e hidrofóbica ao seu sítio ligante.⁹ O sucesso deste tipo de abordagem exige que duas questões sejam previamente resolvidas: inicialmente, deve-se eleger um alvo biológico de relevância comprovada na gênese ou progressão de uma doença. Em segundo lugar, deve-se desenvolver agentes terapêuticos que modulem a atividade do alvo de maneira apropriada e que apresentem níveis tóxicos toleráveis ao paciente.²⁰

Já as estratégias baseadas em ligantes utilizam moléculas orgânicas com atividade biológica conhecida, funcionando como moldes para a triagem em bases de dados de novas entidades químicas com algum nível de similaridade, compartilhando com estes moldes a mesma atividade biológica. Dessa forma, compostos são selecionados de acordo com os mais variados métodos de similaridade molecular e farmacofórica, direcionados por relações entre propriedades estruturais e atividade biológica.²¹⁻²³

Técnicas como Química Combinatória (QC) e ensaios biológicos automatizados em larga escala (“high-throughput screening” - HTS), têm aumentado consideravelmente o número de compostos avaliados quanto à sua atividade. O uso de informações estruturais obtidas por triagem virtual pode reduzir drasticamente esse número, tornando o processo mais dinâmico e otimizado, uma vez que prioriza as moléculas mais interessantes dentre as milhares disponíveis no início do projeto.²⁰

Dessa forma, o objetivo principal de uma triagem virtual é o de identificar os compostos de uma biblioteca que tenham maior probabilidade de se ligar a um alvo biológico. Com os avanços tecnológicos, a cada dia fica mais palpável a realização e aplicação prática dos conceitos de triagem virtual, em particular na otimização de compostos em espaços químicos que sejam acessíveis sinteticamente e que, concomitantemente, estejam de acordo com os padrões farmacofóricos exigidos para o espaço químico de fármacos e substâncias bioativas em Química Medicinal (Figura 1).²⁴⁻²⁶

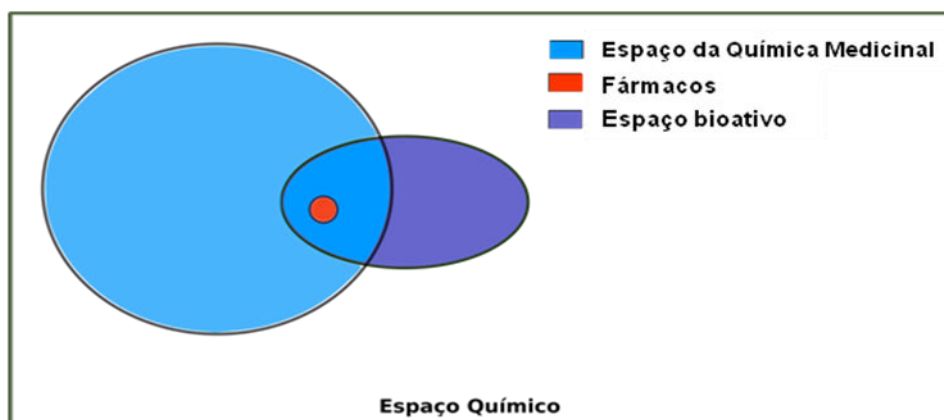


Figura 1. Esquema representando as otimizações nas triagens para a pesquisa farmacêutica, onde o Espaço Químico representa o imenso universo de possíveis moléculas. O espaço bioativo representa todas as moléculas com características farmacofóricas que favoreçam sua interação com algum alvo biológico. Já o espaço de fármacos representa apenas uma pequena porção na interseção do espaço acessível à Química Medicinal com o espaço bioativo. O espaço de fármacos engloba o seleto grupo de substâncias bioativas que são seguras e eficientes no tratamento de processos patológicos. Adaptado de Drewry & Macarron (2010).²⁷

2. Bibliotecas Virtuais

Para a execução do procedimento de triagem virtual, a utilização de uma biblioteca virtual de compostos químicos é uma etapa crucial que deve fornecer compostos com o maior grau possível de diversidade molecular, sejam eles já conhecidos e consagrados por suas atividades biológicas e terapêuticas, sejam compostos já sintetizados e disponíveis para ensaios biológicos, moléculas hipotéticas (desenhadas em computador antes da síntese) ou até produtos de origem natural. A qualidade da base de dados utilizada é de vital importância, pois é a fonte de onde serão selecionados os compostos promissores para futuros ensaios biológicos.²⁸

Conceitos como diversidade e representatividade são utilizados para assegurar uma boa amostragem da biblioteca virtual. A diversidade é um fator que almeja garantir que as moléculas selecionadas sejam diferentes entre si e possam cobrir o maior espaço químico possível de substâncias bioativas (Figura 1), maximizando, assim, a probabilidade de encontrar um conjunto diverso de moléculas promissoras durante uma simulação de triagem virtual. Já a representatividade almeja garantir que as amostras da biblioteca tenham uma distribuição equivalente nas diversas direções do espaço químico, assegurando, assim, que não existam muitas amostras moleculares cobrindo um espaço químico restrito e poucas cobrindo o espaço químico remanescente,²⁸ fato que pode conferir baixo grau de confiabilidade ao experimento por tender a simulações.

Uma biblioteca virtual é composta tipicamente por uma base de dados de compostos disponíveis para aquisição comercial, como é o caso da base de dados ZINC,¹⁶ que já supera 20 milhões de compostos. Como as bases de dados são desenvolvidas para auxiliar a triagem de moléculas com aplicação terapêutica, é muito importante que as propriedades moleculares sejam consideradas. A aplicação de filtros

químicos é uma estratégia usual na redução inicial da biblioteca, com o intuito de priorizar compostos que apresentem propriedades físico-químicas e funcionalidades químicas consistentes com o que se espera para fármacos em geral, ou até para os interesses particulares de uma série de compostos bioativos.²⁹⁻³¹

Nos últimos anos, grande ênfase tem sido direcionada à caracterização de propriedades “drug-like”, incluindo a Regra dos Cinco, de Lipinski,³² com o intuito de minimizar o risco de seleção de moléculas com a atividade biológica desejada, mas que são inadequadas do ponto de vista de propriedades muito importantes em farmacoterapia, tais como: biodisponibilidade, permeabilidade, absorção, metabolismo, excreção e toxicidade. A escolha de uma biblioteca virtual de compostos com propriedades consistentes com o conceito “drug-like” pode reduzir o tempo do ciclo de desenvolvimento e minimizar a inviabilidade de aplicação terapêutica de moléculas bioativas por inadequação de propriedades farmacocinéticas e/ou toxicidade acentuada,^{33,34} propriedades estas que representam o principal gargalo no insucesso de candidatos a fármacos.³⁵⁻³⁷

3. Triagem Virtual Baseada na Estrutura do Alvo Molecular

A estratégia de triagem virtual baseada na estrutura do alvo-molecular está associada à busca de ligantes através de métodos computacionais que consideram a estrutura 3D de um alvo terapêutico.³⁸ O objetivo central é o de prever compostos de uma base de dados capazes de interagir com o sítio ligante do alvo-molecular e ordenar estas moléculas de acordo com a sua afinidade pelo sítio receptor, com o intuito de identificar ligantes promissores com potencial atividade farmacológica. Os cálculos de docagem representam a abordagem central utilizada na triagem baseada em estrutura.³⁹⁻⁴¹

As técnicas de docagem, desenvolvidas para prever a melhor orientação e conformação de uma molécula ligante no seu sítio receptor, são empregadas há algum tempo no processo de planejamento de fármacos. A ideia geral é obter um conjunto de conformações do complexo ligante-receptor e classificá-las em ordem de prioridade com base em suas estabilidades energéticas.^{10,42}

A capacidade de predição dos modos de ligação e afinidade de compostos que interagem com o sítio ligante de uma proteína é um dos maiores desafios do desenvolvimento de fármacos baseado em estrutura. Consequentemente, o número de

algoritmos desenvolvidos para programas de docagem é amplo. A capacidade de reproduzir a orientação experimental de ligantes em relação aos modos de ligação observados em complexos com estrutura já resolvida é a característica mais importante do cálculo de docagem. A reprodução de um modo de ligação experimental normalmente é considerada bem-sucedida quando o desvio médio quadrático (“Root Mean Square Deviation”, RMSD) calculado se encontra abaixo de 2,0 Å (Figura 2) entre a orientação cristalográfica (experimental) e a orientação obtida durante o cálculo.⁴³

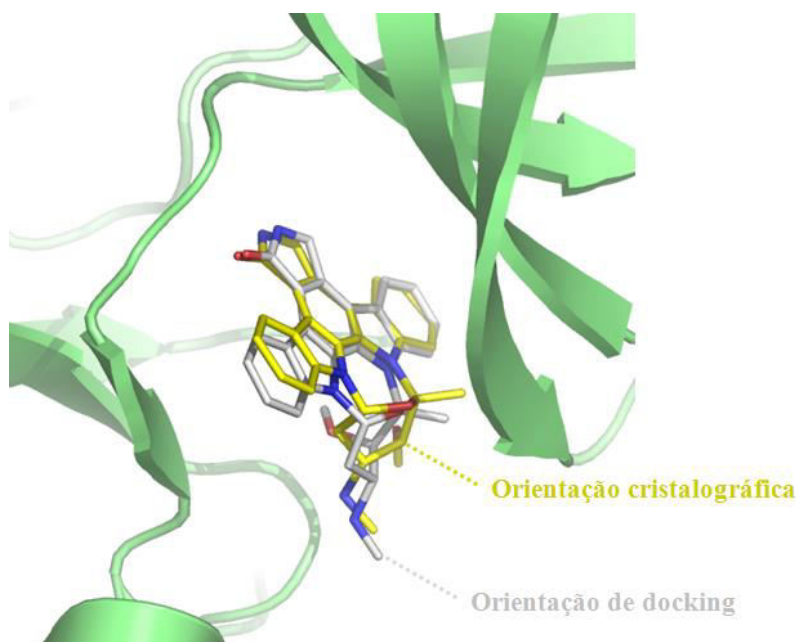


Figura 2. Reprodução do modo de ligação experimental (orientação cristalográfica) da estaurosporina no sítio ligante da CDK2 através do cálculo de docagem, com RMSD de 0,91 Å. Simulação realizada com programa GOLD 3.1.1.⁴³ Figura gerada com o programa PyMOL.⁴⁴

Vários programas de “docagem” que usam diferentes filosofias metodológicas estão disponíveis. Alguns dos programas mais utilizados na atualidade incluem: FlexX,⁴⁵ DOCK,⁴⁶ GOLD⁴³, GLIDE^{47,48} e AutoDock.⁴⁹ DOCK e AutoDock têm a vantagem adicional de serem gratuitos para instituições acadêmicas. A técnica de docagem é popular justamente pelo seu potencial de identificação de compostos com

características químicas e estruturais distintas de ligantes usuais e conhecidos para um alvo molecular específico,⁵⁰ principalmente quando a base de dados utilizada garante elevado grau de diversidade e representatividade molecular, permitindo, assim, explorar ao máximo o grau de variabilidade química requerido para interação de possíveis ligantes em um alvo-

molecular através da complementaridade com o sítio receptor.

Os protocolos de docagem são formados, de maneira geral, por dois componentes: o algoritmo de busca e a função de escore. O algoritmo é utilizado na busca sistemática de orientações e conformações do ligante no sítio receptor e a função de escore é aplicada aos possíveis modos de ligação propostos pelo algoritmo, com o propósito de classificá-los por afinidade de acordo com suas interações com os resíduos de aminoácidos do sítio receptor.^{42,51}

3.1. Algoritmos de busca

Os algoritmos de busca, genericamente, são um conjunto de regras e parâmetros aplicados com o objetivo de explorar a flexibilidade da estrutura dos ligantes em simulações de docagem. Para o êxito da triagem virtual, é imprescindível o uso de algoritmos eficientes que não apenas encaixem cada molécula do ligante realisticamente no sítio receptor, mas que também obtenham conformações energeticamente viáveis do mesmo. Para cada molécula, os algoritmos de busca tendem a criar várias orientações diferentes, com distintos modos de ligação no sítio receptor.^{25,52} Embora cada uma dessas tentativas pareça ser aleatória e independente, um grande número de algoritmos adota uma heurística baseada na química ou geometria dos átomos envolvidos, como os algoritmos dos programas FlexX⁵³ e DOCK. Em outros casos, podem adotar uma técnica de otimização padrão, como a estratégia de algoritmo genético implementada nos programas AutoDock⁵⁴⁻⁵⁶ e GOLD.⁴³

Nas estratégias pioneiras de docagem, os algoritmos empregados mantinham o ligante e a proteína em sua conformação inicial durante todo o processo, considerando-os como entidades não flexíveis no espaço conformacional (docagem de “corpo rígido”), reduzindo assim a demanda computacional

que seria empregada em uma busca flexível do complexo ligante-proteína.^{51,57}

O maior problema desta abordagem é que a flexibilidade do ligante não é considerada. Na maioria das vezes, moléculas pequenas possuem grande variação conformacional, com diversos estados de baixa energia. Até mesmo nas que têm poucas ligações rotacionáveis podem ocorrer diferenças significativas entre a conformação experimental e as conformações sugeridas pelo método de docagem.^{53,57,58} Na tentativa de suprir esta deficiência, os algoritmos tornaram-se mais elaborados, permitindo uma busca mais completa no espaço químico, com processamento de cálculo adequado a cada situação, incluindo a consideração de propriedades físico-químicas e da flexibilidade do ligante. Essas melhorias visaram, principalmente, considerar aspectos do ajuste induzido do complexo ligante-receptor e descartar conformações energeticamente desfavoráveis do ligante,⁵⁷ o que por um lado contribui para o aumento do gasto computacional necessário para a realização do procedimento, mas incrementa substancialmente a precisão do método e a qualidade do resultado final.

Os algoritmos de busca, quanto à análise da flexibilidade do ligante, podem ser agrupados em três categorias: sistemáticos, randômicos ou baseados em simulação (Tabela 1), sendo que a maioria dos programas tende a mesclar algumas destas abordagens.^{19,29,59}

Os algoritmos randômicos operam realizando mudanças aleatórias, geralmente alterando grau de liberdade (ligações rotacionáveis) de uma molécula, explorando, assim, diversas conformações possíveis. Nesta categoria destacam-se as simulações de Monte Carlo, tal como no programa MOE-Dock,⁶⁰ e os algoritmos genéticos, representados, principalmente, pelos programas AutoDock⁶⁰ e GOLD [Tabela 1].^{5,19,60}

Na estratégia de Monte Carlo, mudanças aleatórias são aplicadas aos movimentos translacionais, rotacionais e torcionais da

molécula. Após cada movimento, o complexo é minimizado e a energia da nova estrutura calculada. A aceitação de um novo modo de ligação é baseada em sua energia: se for mais

baixa, a configuração é aceita; se for mais alta, a configuração é aceita, com certa limitação, que depende da magnitude desse aumento.¹⁹

Tabela 1. Principais algoritmos utilizados nas estratégias de docagem

Algoritmos de Busca	Estratégia	Programa de Docagem
Randômicos	Algoritmo Genético	AutoDock GOLD DARWIN
	Monte Carlo	MOE-Dock ICM MC DOCK DOCK VISION AFFINITY QXP GLIDE
Sistemáticos	Incremental	DOCK FlexX Glide Surflex-Dock Hammerhead HOOK
	Base de dados de confôrmeros	Flexibase/FLOG
	Não-estocástico	FRED
Simulação	Dinâmica molecular	DynaDock
	Minimização de energia	DOCK

Os algoritmos genéticos, também conhecidos como evolucionários, como o próprio nome sugere, são programados para usar o mesmo princípio da evolução biológica.^{25,52} Geralmente, diz-se que os algoritmos genéticos pertencem à classe dos algoritmos não determinísticos, isto é, o tipo de cálculo usado prevê uma tomada de decisão a partir de sua heurística, fornecendo soluções satisfatórias para a otimização combinatória a um custo acessível, com o propósito de encontrar soluções, que por exemplo, no caso da docagem, buscariam prever a conformação mais estável.^{51,62} O programa GOLD, por exemplo, utiliza uma variação desse algoritmo, com modificações de diedros dos ligantes, geometria de anéis e

modificações na posição do ligante no sítio ativo do alvo molecular.⁴³

Os algoritmos genéticos são representados por uma série binária que apresenta uma estrutura de codificação semelhante ao DNA.⁵² Esta semelhança é estratégica, pois permite simular operações genéticas, como mutações e recombinações (Figura 3A). A conformação de uma molécula deve ser representada de modo a permitir que um processo evolucionário de mutação e seleção ocorra. De forma prática, todas as informações de ângulos de torção de ligações de uma molécula são armazenados como uma sequência numérica, que irá corresponder ao “cromossomo”. Dentro deste “cromossomo”, cada “gene”

corresponde a um ângulo de torção (Figura 3B).²⁵

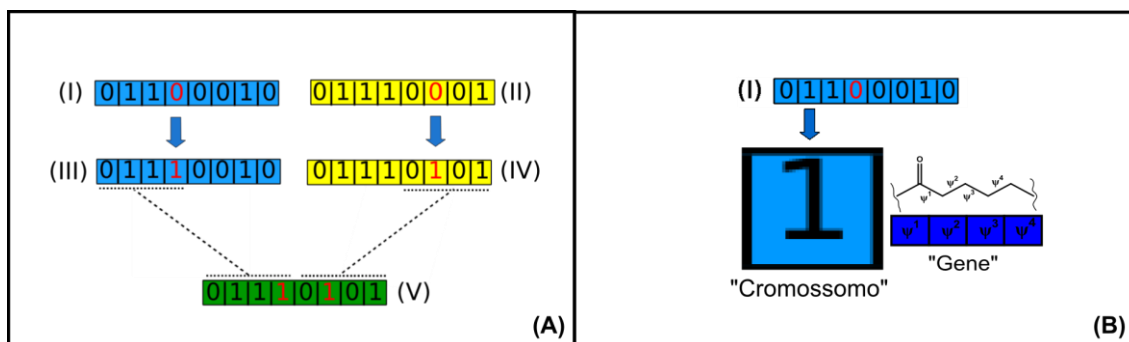


Figura 3. A) Exemplo de mutações e recombinação em sequências binárias de 8 bits de extensão. Os bits em vermelho destacam as mutações, que geram os indivíduos III e IV. O indivíduo V é formado através da recombinação entre III e IV. **B)** Cada valor numérico que representa um indivíduo, possui subgrupos de números responsáveis por armazenar informações relativas a ângulos de rotações das partes da molécula. Adaptado de Patrick (2009).²⁵

Os algoritmos sistemáticos exploram os graus de liberdade das moléculas, normalmente, através de sua construção incremental no sítio receptor. Os algoritmos incrementais, como o próprio nome sugere, utilizam-se do método de construção incremental através da fragmentação inicial da molécula. A simulação se inicia com o posicionamento de um fragmento do ligante, denominado âncora, no sítio ativo do alvo molecular. Então, os fragmentos remanescentes dos ligantes são adicionados sucessivamente, até atingir a reconstrução total da molécula com seu posicionamento em relação à estrutura do receptor. Assim, um algoritmo de construção incremental possui três fases: a seleção do fragmento âncora, o seu posicionamento no sítio e a fase da construção incremental.^{57,63}

A heurística dos algoritmos incrementais considera o fato de que a maioria das moléculas possui pelo menos um fragmento rígido capaz de interagir com o sítio receptor, como por exemplo, um anel aromático. Esses fragmentos são dispostos em várias posições favoráveis do sítio ativo. Em sequência, soluções de docagem são criadas, baseadas nas posições iniciais destes fragmentos, seguido pelo processo de construção incremental, onde o espaço conformacional

dos novos fragmentos adicionados será explorado.^{19,64} Um exemplo do emprego de algoritmos incrementais pode ser observado com o programa FlexX (Tabela1).⁴⁵

Outra estratégia de busca sistemática é a utilização de bibliotecas pré-concebidas de conformêros. Um exemplo de algoritmo baseado na análise das conformações é o Flexibase/FLOG, desenvolvido por Miller e Kearsley.⁵⁷ O Flexibase⁶⁵ armazena uma pequena série de diferentes conformações para cada molécula de uma determinada base de dados. Então, aproximadamente 25 conformações de cada molécula são selecionadas adotando como critério o desvio médio quadrático entre elas. Por fim, cada conformação da molécula é submetida ao procedimento de docagem utilizando o algoritmo FLOG, o qual considerará cada conformação como estática durante a simulação.^{57,66-69}

Outro programa bastante empregado nas estratégias de docagem é o FRED, que utiliza-se de multiconfôrmeros provenientes de uma biblioteca, além de um arquivo com informações do receptor ("input"). Como resultado final, cria um arquivo final contendo as moléculas com maior probabilidade de se ligarem ao receptor ("output"). Para isso, examina

exaustivamente as possíveis conformações do ligante dentro do sítio ativo, filtrando por meio de complementaridades de forma e dos grupos farmacofóricos antes da classificação. Esta busca exaustiva baseia-se em rotações e translações rígidas de cada confômero, evitando os problemas comuns associados às rotações flexíveis.^{70,71} Como resultado final, FRED retorna uma única molécula "docada" para cada uma das moléculas iniciais, adotando em seu algoritmo os conceitos sistemático e não-estocástico,^{70,72} isto é, realiza a construção incremental no sítio receptor, porém adota como padrão o uso de estruturas com conformações inalteradas (fixas).

Já o programa Surflex-Dock, um dos módulos do pacote computacional SYBYL, da empresa Tripos,⁷³ adota uma estratégia de docagem que combina uma função de escore empírica (semelhante a do Hammerhead)⁷⁴ com métodos de similaridade morfológica,^{75,76} gerando conformações aleatórias de fragmentos do ligante. Além disso, também implementa uma nova forma, mais rápida e precisa de reunir estes fragmentos.^{76,77} O método como estes fragmentos são agrupados remete à abordagem utilizada nos algoritmos genéticos,⁷⁸ entretanto, sua essência é determinística, pois o algoritmo irá seguir um padrão já estabelecido, sem a tomada de decisões que ocorreria caso utilizasse a estratégia dos algoritmos genéticos.⁷⁶

Nos algoritmos de simulação, a estratégia mais empregada é a dinâmica molecular. Para uma análise mais apurada, que obtenha detalhes em nível atômico das interações proteína-ligante, foram desenvolvidos mecanismos de dinâmica molecular que permitem uma melhor exploração da energia livre do sistema, como o aumento da temperatura da simulação. Inclusive, a princípio, todos os graus de liberdade do receptor também podem ser explorados.¹⁹ Um exemplo de aplicação dessa estratégia é o programa DynaDock.⁷⁹

Poucos são os métodos que utilizam simulações de dinâmica molecular no estudo

de docagem proteína-ligante. Isto ocorre porque a demanda computacional é altíssima, e apesar de garantir a flexibilidade total do sistema o grande fator limitante é o tempo gasto para execução das simulações.⁸⁰ De forma geral, o elevado número de moléculas de uma base de dados exige que os cálculos de triagem virtual sejam rápidos e suficientemente precisos para processar o grande número de estruturas envolvidas em um tempo razoável. Entretanto, ao aumentar a velocidade na qual um algoritmo opera, invariavelmente é necessário diminuir a precisão do cálculo.^{25,52,81} Como resultado, essa é uma área de intensa pesquisa, alimentada pela atual necessidade de desenvolvimento de algoritmos mais completos e otimizados.

3.2. Funções de escore

As funções de escore são utilizadas com o intuito de estimar a afinidade de uma dada posição (orientação) do ligante no sítio receptor.⁸² As posições do ligante em relação ao sítio receptor são obtidas em etapas anteriores, através dos algoritmos de busca, como explicitado acima. Embora os métodos de docagem sejam úteis no desenvolvimento de protótipos na busca por novos fármacos, um dos grandes problemas ainda a serem resolvidos, ou melhorados, é a predição das energias de ligação.³⁸

Nos cálculos de docagem, as funções de escore desempenham duas tarefas principais durante as etapas de triagem virtual. Primeiramente, são usadas para diferenciar e avaliar as diversas conformações e orientações que um mesmo ligante pode adotar no sítio receptor. Em seguida, após a realização de simulações com diversos ligantes contidos em bases de dados, a função de escore é aplicada para estimar a afinidade de ligação dos diversos complexos ligante-receptor estudados e, dessa forma, classificá-los de acordo com sua afinidade,^{43,83} facilitando assim a diferenciação entre ligantes promissores e moléculas não-ligantes.

As principais funções de escore implementadas nos mais diversos programas disponíveis podem ser classificadas, genericamente, em empíricas, em baseadas em campos de força de mecânica molecular e em baseadas em conhecimento (“knowledge-based”). As funções empíricas são calibradas em relação a um conjunto de dados de afinidade de complexos ligante-receptor obtidos experimentalmente. Nas funções baseadas em mecânica molecular, a afinidade de ligação é estimada de acordo com o somatório das energias de interação de van der Waals e eletrostática entre os átomos do ligante e do sítio receptor. As funções de escore do tipo “knowledge-based” são derivadas de análises de preferências estatísticas das distâncias de interação entre os diferentes átomos baseadas no conhecimento de complexos ligante-proteína com estrutura resolvida.^{38,84} Geralmente, contatos atômicos que ocorrem com frequência maior do que a média são energeticamente favoráveis. As funções “knowledge-based” são somas de várias contribuições desses contatos para as combinações de átomos da proteína e do ligante.⁸³

Os métodos de docagem utilizam as funções de escore, basicamente, de duas maneiras. A primeira abordagem lança mão da função de escore completa para ordenar uma conformação do ligante em relação ao receptor. Então, o sistema é modificado pelo algoritmo de busca, e a nova conformação obtida é classificada pela mesma função de escore. O método alternativo utiliza a função de escore em dois estágios. Nessa abordagem, uma função reduzida é aplicada ao sistema com o objetivo de direcionar o algoritmo de busca para regiões do sítio receptor em que o escore de interação é mais alto e, então, uma função de escore mais refinada ranqueia as conformações obtidas nas diferentes posições. A função de escore reduzida, normalmente, considera apenas poucos tipos de interação, como por exemplo, ligações de hidrogênio, omitindo outros termos importantes que serão avaliados pela função de escore refinada.⁴² Portanto, dependendo das condições como o

tamanho da base de compostos a ser analisada ou a potência dos computadores disponíveis pode-se optar por uma ou outra metodologia, sendo que a função de escore completa é mais robusta, porém demanda mais tempo.

3.3. Conformação do alvo molecular

A conformação do alvo molecular proteico é um fator importante na obtenção de resultados confiáveis em cálculos de docagem, pois a maior parte dos métodos trabalha com amostras conformacionais estáticas do alvo molecular. O nível de exatidão é afetado sobremaneira pela qualidade das estruturas proteicas. Considerando as conformações *holo* (conformação do receptor complexado ao ligante), *apo* (conformação do receptor livre do ligante) e outras obtidas por modelagem molecular para um mesmo receptor proteico, normalmente, os melhores resultados são obtidos com a conformação *holo*, seguidos pela *apo* e, posteriormente, estruturas modeladas. Isso se deve ao fato de a conformação *holo* ser influenciada pela presença de um ligante através de um ajuste induzido, diferentemente de uma conformação proteica sem a presença de qualquer ligante (*apo*) que possa interferir em sua conformação. Apesar dessa regra geral, algumas exceções podem ocorrer. Em alguns casos, a conformação *holo* só permite bons resultados de simulação com moléculas da base de dados estruturalmente semelhantes ao ligante presente na estrutura resolvida do complexo ligante-receptor, perdendo qualidade em simulações com moléculas com características estruturais distintas e um padrão diferente de ligação.^{10,23,85}

A conformação *apo* pode ser inadequada para a acomodação do ligante, pois sítios ativos que apresentam mudanças conformacionais podem ter resíduos em posições que não traduziriam um estado adequado para interagir corretamente com o ligante estudado e, dessa forma, dificultar a

obtenção de resultados confiáveis em procedimentos de docagem. As estruturas modeladas, por sua vez, dependendo de sua qualidade podem apresentar resíduos ou suas cadeias laterais em locais impróprios, fato que também dificulta a exatidão do procedimento.¹⁰

É devido a estes fatores que a preparação do receptor proteico antes da realização das simulações de docagem é tão importante. Uma das alternativas é considerar a flexibilidade da estrutura do receptor e permitir mudanças conformacionais durante a simulação, pois os alvos-moleculares em sistemas biológicos estão em contínuo movimento.⁸⁶ Apesar das estruturas resolvidas por difração de raios-X serem comumente empregadas no planejamento de substância bioativas, vale ressaltar que a maioria das informações estruturais acerca de proteínas depositadas no PDB é limitada à conformação mais estável quando cristalizada em condições artificiais. Em condições fisiológicas um inibidor interage com conformações variadas da proteína e com maior afinidade quando comparado com a conformação sugerida por cristalografia de raios-X, por exemplo.⁸⁷

A não consideração de flexibilidade em um cálculo de docagem pode acarretar em pouca exatidão nas predições de modos de ligação, redução na eficiência de seleção de moléculas ativas durante a triagem virtual e baixa correlação entre score e afinidade de ligação. Quando a flexibilidade do receptor é incorporada às simulações, os riscos associados à obtenção de conformações inadequadas da proteína para o sistema estudado são minimizados.¹⁰ Alguns dos programas mais recentes de docagem permitem a flexibilidade da estrutura do receptor, pelo menos parcial, ficando restrita às cadeias laterais do sítio ligante através da aplicação de uma biblioteca de rotâmeros.

Embora a flexibilização das cadeias laterais nos cálculos de docagem requeira maior gasto computacional, ela permite movimentos localizados da proteína resultando em um melhor encaixe do ligante

ao seu sítio receptor.⁸⁸ Normalmente, apenas algumas cadeias laterais dos resíduos do sítio ligante, selecionadas pelo usuário dos programas de acordo com o seu conhecimento do sistema estudado, são flexíveis durante a simulação. No entanto, estas abordagens ainda são incipientes e desconsideram maiores rearranjos conformacionais na proteína, como movimentos no esqueleto peptídico, por exemplo.⁸⁹

Uma das alternativas consiste em mesclar várias conformações da estrutura do alvo molecular em uma única conformação média, para considerar conformações de resíduos que poderiam se mostrar essenciais na formação do complexo ligante-receptor. Uma estrutura média final pode ser gerada através da média das coordenadas atômicas das várias conformações envolvidas, fato que pode levar à obtenção de conformações artificiais, que podem ser muito distintas das conformações biológicas, sobretudo em relação à região que compõe o sítio de ligação do alvo molecular.⁹⁰ Um protocolo mais robusto consistiria em simular o encaixe do ligante em seu respectivo sítio de ligação através de docagem automatizada utilizando múltiplas conformações do receptor, e não somente uma média, processo também conhecido como “ensemble docking”, o qual requer um gasto computacional maior quando comparado à triagem com apenas uma conformação. As múltiplas conformações mimetizam o equilíbrio conformacional que caracteriza o estado natural de um alvo molecular proteico, fornecendo, assim, um modelo que considera um maior grau de liberdade estrutural. Em relação ao procedimento de múltiplas conformações (“ensemble docking”), o conjunto de conformações da proteína empregado no cálculo pode advir de conformações obtidas experimentalmente, como por exemplo, de um conjunto de complexos do mesmo alvo molecular com ligantes distintos depositados no PDB, ou mesmo através de modelos gerados por métodos computacionais.^{89,91}

4. Triagem Virtual Baseada em Ligantes

A triagem virtual baseada em ligantes é uma estratégia que utiliza compostos com atividade biológica já conhecida como fonte de informação inicial, sem levar em conta a estrutura de um alvo molecular. Esta abordagem visa a identificação de moléculas com certa semelhança estrutural e que possam compartilhar também atividade biológica.⁹² O raciocínio central se baseia no princípio de que as moléculas que compartilham algumas semelhanças estruturais entre si podem apresentar também atividade biológica semelhante. Em geral, os métodos utilizam um conjunto de ligantes ativos já conhecidos como referência para extrair compostos estruturalmente semelhantes a partir das bibliotecas virtuais de compostos. A triagem virtual baseada em ligantes é uma estratégia muito útil, principalmente, quando nenhuma ou pouca informação advinda da estrutura 3D do alvo molecular estudado é conhecida.⁹³ Nessa abordagem, a busca de novos ligantes fica restrita à exploração do espaço químico oferecida pelos compostos ativos conhecidos. Através da sobreposição dos compostos ativos, por exemplo, é possível criar um esboço das principais características compartilhadas entre as moléculas no espaço tridimensional, como possíveis grupos capazes de realizar ligações de hidrogênio, bem como grupos hidrofóbicos, com carga positiva ou negativa.⁹⁴⁻⁹⁶

Inicialmente, a triagem virtual baseada em ligantes foi desenvolvida a partir de descritores 2D ou características moleculares específicas (“fingerprints”) derivadas da estrutura de compostos ativos de referência e comparados com os correspondentes descritores das moléculas das bases de dados, lançando mão de uma medida de similaridade, como por exemplo o famoso coeficiente de Tanimoto. Este método é, geralmente, muito rápido e fornece como resultado uma lista de compostos classificados de acordo com a sua

similaridade ao quimiotipo da(s) molécula(s) de referência. Posteriormente, buscando aumentar a diversidade estrutural dos compostos protótipos (“hits”) fornecidos, foram desenvolvidos diferentes métodos baseados em descritores 3D mais elaborados.⁹⁷

Dentre os métodos baseados na geometria 3D, alguns são dependentes do alinhamento de conformações moleculares. Talvez o método mais amplamente empregado exigindo representações 3D, é aquele que explora o conceito de similaridade aos grupos da molécula essenciais à atividade (farmacóforo).^{93,98,99}

O sucesso das técnicas de triagem virtual baseada em ligantes depende, portanto, fortemente da natureza das relações estrutura-atividade. Estas fornecem um conjunto de restrições e parâmetros que podem ser de grande valia na busca de novas estruturas químicas com, ao menos, níveis de atividades biológicas similares aos das moléculas bioativas já conhecidas.¹⁰⁰ Estes métodos fornecem enriquecimentos significativos sobre a seleção aleatória de moléculas em bases de dados. Assim, as moléculas de maior pontuação podem ser priorizadas em testes experimentais, reduzindo a relação custo/benefício nos programas de desenvolvimento de fármacos.⁹⁸

O crescimento cada vez maior das bases de dados de pequenas moléculas bioativas, o desenvolvimento de novos algoritmos, melhorias em componentes de computadores (“hardwares”), programas (“softwares”) e métodos de triagem têm sido cada vez mais aplicados no cotidiano da identificação de compostos ativos, permitindo o contínuo aprimoramento desta estratégia e ampliando o campo da triagem virtual na interface químico-farmacêutica.^{91,101}

4.1. Triagem virtual por padrão farmacofórico

Uma abordagem pertencente ao grupo de técnicas de triagem virtual baseada em ligantes é a busca através da determinação do padrão farmacofórico. A modelagem do farmacóforo é uma abordagem amplamente empregada para um conjunto de ligantes ativos conhecidos, cujos grupos funcionais responsáveis por interações específicas fármaco-receptor são analisados através de alinhamentos no espaço tridimensional.^{102,103} O conceito de farmacóforo foi primeiramente introduzido por Paul Ehrlich, em 1909, como "uma estrutura molecular (phóros) que carrega as características essenciais, responsáveis pela ação biológica de um fármaco (pharmacon's)". O farmacóforo inclui todas as principais características envolvidas na interação do ligante com o sítio receptor. Através desse conhecimento alguns grupos químicos podem ser incorporados a determinados ligantes a fim de elevar a especificidade, visando otimizar sua atividade biológica.¹⁰⁴ Na prática, o farmacóforo engloba características químicas específicas, tais como potencialidade para realização de ligações de hidrogênio, interações eletrostáticas e hidrofóbicas, entre outras.¹⁰⁵ Dessa forma, um modelo farmacofórico sugerido para uma classe de ligantes pode ser utilizado como molde na busca e seleção de novos protótipos que satisfaçam os principais requisitos moleculares responsáveis pela atividade biológica em determinado alvo molecular.

Os elementos farmacofóricos chaves em uma busca podem ser: um grupo específico de átomos, uma região molecular volumosa, grupamentos doadores e aceptores de ligação de hidrogênio, grupamentos com cargas, hidrofóbicos e/ou anéis aromáticos, associados a restrições geométricas incluindo distâncias e ângulos de ligação. Assim, um grupo farmacofórico acceptor de ligação de hidrogênio, por exemplo, permite a busca de diversos padrões estruturais, como moléculas contendo grupo hidroxila, carbonila ou um

grupo éster, todos passíveis de realizarem interações por ligação de hidrogênio.¹⁰² Esses elementos se tornam mais representativos quando as principais interações de um grupo de ligantes com o seu alvo molecular são bem estabelecidas. Dessa forma, a busca pode ser direcionada para potenciais ligantes que apresentem similaridades farmacofóricas relevantes a um conjunto de compostos bioativos, e não apenas para similaridades comuns, ao acaso, de grupamentos funcionais que mesmo sendo comuns ao grupo de ligantes não são importantes na interação com o receptor biológico. Ainda assim, essa abordagem apresenta limitações, já que nem sempre se dispõe de estruturas cristalográficas dos compostos ativos conhecidos complexados com seu alvo biológico, e mesmo quando há, a conformação cristalográfica do ligante geralmente varia em relação a sua real conformação bioativa, já que aquela não é obtida sob condições fisiológicas.^{87,89}

Outra limitação encontrada quando se trabalha com modelos farmacofóricos está associada à incerteza em relação a configuração espacial dos grupos farmacofóricos. Configurações específicas demais levam a seleção de compostos com diversidade estrutural muito pobre, e por outro lado, configurações mais gerais levam a um modelo farmacofórico que gera um grande número de resultados falso-positivos.⁸⁹

Um modelo farmacofórico 3D não é transferível para outro modelo 3D obtido para um conjunto diferente de ligantes, ou seja, um mesmo alvo biológico pode possuir diferentes moléculas bioativas que apresentam diferentes conformações e orientações em seu sítio ativo. Assim, o padrão farmacofórico irá depender do grupo de moléculas selecionadas para sua derivação.¹⁰⁶ Para criar um modelo farmacofórico teórico, geralmente, o espaço conformacional de cada ligante é explorado e enumerado em vários confôrmeros. Uma vez gerado, cada confôrmero será tratado como um corpo rígido durante o mapeamento do

farmacóforo. Nem sempre é necessário que as coordenadas de uma característica molecular para uma molécula no espaço 3D se encaixe perfeitamente com o farmacóforo correspondente comum a todo o conjunto de ligantes estudado, pois, normalmente, existe uma margem de tolerância definida pelo usuário, a qual forma uma zona de alcance em torno de cada característica do farmacóforo. Esta zona de alcance é influenciada pela incerteza provocada pela flexibilidade e adaptabilidade do sistema proteína-ligante. Um raio da zona de alcance maior indica uma maior incerteza associada ao grupo funcional correspondente.¹⁰⁴

Chen e colaboradores¹⁰⁷ compararam métodos de triagem virtual baseados em docagem e farmacóforo através da análise de 16 conjuntos de buscas (8 alvos moleculares, cada um testado com duas bases de dados). A estratégia baseada em farmacóforo mostrou um maior fator de enriquecimento quando comparada à estratégia de docagem, inclusive com níveis superiores de recuperação de moléculas bioativas.

Os inibidores da enzima HMG-CoA

redutase, por exemplo, utilizados na terapia da hiperlipoproteinemia, apresentam um farmacóforo definido, o qual pode ser analisado após a sobreposição dos modos de complexação dos ligantes presentes nos complexos com a enzima depositados no PDB, ou mesmo das conformações de baixa energia geradas por um programa capaz de derivar um padrão farmacofórico. Assim, os inibidores da HMG-CoA redutase apresentam em comum um núcleo central e outro periférico, compostos por anéis aromáticos, além de uma cadeia alifática hidroxiácida com quatro grupamentos aceptores e/ou doadores de ligação de hidrogênio e um grupo ionizável aniônico. O anel central aromático também apresenta substituintes de natureza hidrofóbica (Figura 4A). A análise visual dos complexos inibidores-HMG-CoA redutase¹⁰⁸ revela que todos estes grupamentos desempenham algum papel na interação com o sítio ligante da enzima (Figura 4B), embora outros grupamentos não compartilhados entre a série de inibidores também sejam importantes para cada inibidor, o que confere a diferença de potência observada entre os mesmos.

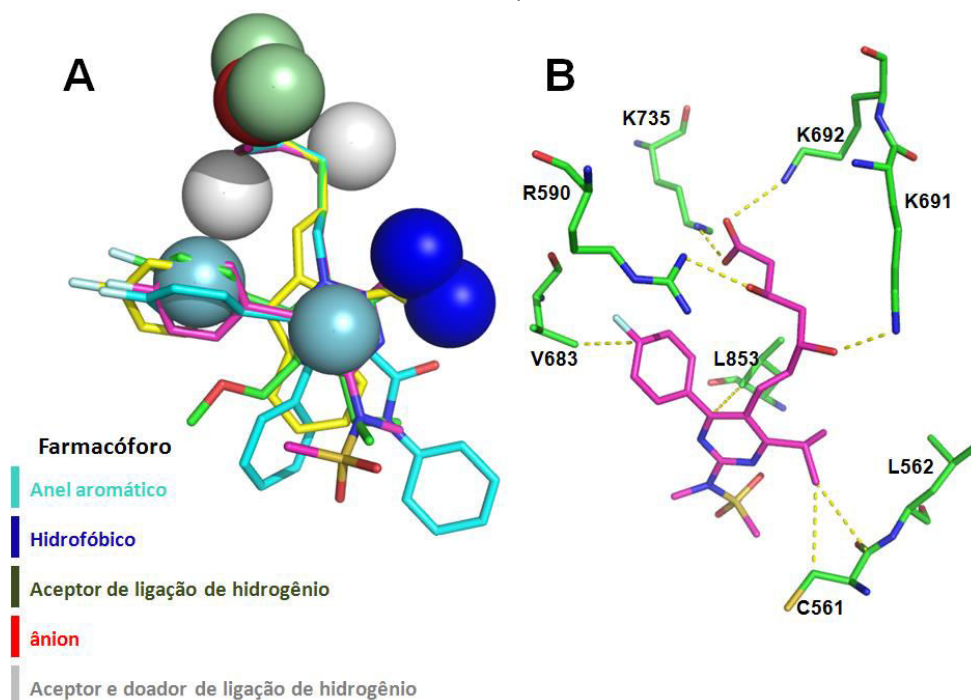


Figura 4. (A) Modelo farmacofórico gerado para inibidores da HMG-CoA redutase com o programa Pharmagist,⁹⁹ entre eles a rosuvastatina, cerivastatina, fluvastatina e atorvastatina. (B) Interações dos grupamentos farmacofóricos presentes na estrutura da rosuvastatina com os resíduos da HMG-CoA redutase. Figura gerada com o programa PyMOL⁴⁴

O conjunto de características químicas definido acima (Figura 4) e compartilhado pela série de inibidores seria o requerimento mínimo esperado para a obtenção de compostos com atividade inibitória relevante frente à HMG-CoA redutase. Dessa forma, o padrão farmacofórico abstraído para estes inibidores confere informação importante na busca de novos inibidores que mantenham as mesmas características químicas no espaço 3D, uma vez que tal modelo poderia ser utilizado como molde na seleção de novas moléculas em uma triagem virtual. Vale ressaltar que a natureza de cada componente do grupo farmacofórico é mais importante que o próprio grupo químico presente na série original de compostos, uma vez que grupamentos químicos distintos, entretanto com características químicas semelhantes, podem desempenhar a mesma função na manutenção de interações com o alvo molecular. Um exemplo prático pode ser evidenciado pelos grupamentos hidroxila e carbonila, ambos capazes de atuar como aceptores de ligação de hidrogênio.²⁵

Outra abordagem de triagem virtual baseada em farmacóforo consiste em utilizar um banco de dados de farmacóforos derivados de inibidores de alvos moleculares de interesse farmacêutico, como o PharmTargetDB, e assim, realizar uma triagem reversa, ou seja, uma molécula em questão é analisada frente a diversos alvos. PharmTargetDB pertence ao servidor online gratuito PhamMapper¹¹⁰ e contém mais de 7000 modelos farmacofóricos, sendo que cada modelo foi gerado a partir de um único inibidor, a partir de sua conformação obtida de bancos de dados que contém complexos proteína-ligante, como o TargetBank, BindingDB e DrugBank. A expressão “reversa” para triagem virtual se aplica quando ao invés de buscar moléculas para um alvo molecular específico, como qualquer triagem convencional, o método de triagem reversa busca possíveis alvos moleculares para uma molécula de interesse. Nesse caso, uma lista de compatibilidades entre as estruturas que melhor se encaixam em cada modelo farmacofórico da base de dados pode ser

obtida e utilizada para relacionar cada composto filtrado com uma ação específica em determinado alvo molecular. Assim como na triagem convencional, este método de triagem reversa também apresenta limitações relacionadas com as disparidades entre as conformações das estruturas dos inibidores complexados existentes, usadas para gerar o modelo farmacofórico, e a sua real conformação bioativa.^{86,89} As aplicações deste método permitem a caracterização de possíveis propriedades farmacológicas inerentes a novos compostos já sintetizados e também a sugestão de um possível espectro de atividade farmacológica para fármacos consagrados e compostos bioativos já conhecidos. A aplicação dessa estratégia permite também a triagem computacional de potenciais sítios de ação capazes de associar os compostos filtrados a efeitos tóxicos e curso metabólico, através do alinhamento desses compostos a um modelo farmacofórico específico para ligantes de proteínas da família do citocromo P450, por exemplo, se alinhando a um novo paradigma estabelecido nas etapas iniciais do processo de desenvolvimento de fármacos, a preocupação com propriedades farmacocinéticas adequadas e níveis toleráveis de toxicidade.^{19,109,110}

4.2. Triagem por similaridade

Outra estratégia comumente aplicada durante o processo de triagem virtual baseada em ligantes consiste na busca por similaridade. O princípio da propriedade de similaridade preconiza que moléculas estruturalmente similares possuem alta probabilidade de compartilharem também propriedades físico-químicas e atividades biológicas semelhantes.¹¹¹ Entretanto, a similaridade e dissimilaridade entre estruturas químicas não são assim objetivamente definidas. Existem várias relações entre a estrutura e a atividade que demonstram que compostos similares podem ter atividades biológicas diferentes¹¹².

Kubinyi, H. (2002)¹¹², em seu trabalho sobre a similaridade química e atividades biológicas afirma que simples mudanças bioisostéricas de átomos e grupos em alguns compostos biologicamente ativos podem resultar em efeitos inesperados, alguns desses efeitos incluem até novos modos de ação. O que deve ser levado em consideração sobre a similaridade e a diversidade de moléculas é a sua dependência em relação à estrutura alvo tridimensional, assim como as propriedades do sítio de ligação do alvo biológico, ou seja, sua complexação ao receptor biológico.

As busca por similaridade consiste em uma das estratégias de triagem mais simples e com menor gasto computacional. Na aplicação desta estratégia, como em qualquer busca baseada em ligantes, o ponto de partida corresponde à seleção de uma ou mais moléculas com atividade biológica conhecida (denominadas estruturas-alvo ou de referência). Através da triagem em bases de dados de moléculas virtuais é possível filtrar moléculas que são estruturalmente mais similares às de referência. A lista de moléculas resultante da triagem indica as moléculas mais prováveis de exibir propriedades similares às da(s) estrutura(s) de referência, tornando-as candidatas para

testes biológicos. Assim, se o princípio de similaridade é aplicável, um conjunto de moléculas que ainda não foi testado para atividade biológica, mas que possui similaridade estrutural a uma ou mais moléculas de referência, apresenta, portanto, alta probabilidade de fornecer substâncias também bioativas.¹¹³

Do ponto de vista prático, para efetuar a comparação entre duas moléculas é necessário empregar um padrão de medida, também conhecido como coeficiente de similaridade, o qual quantificará o grau de semelhança entre as entidades químicas avaliadas. Existem inúmeros coeficientes de medida (Tabela 2), sendo que o mais empregado em triagens virtuais é o coeficiente de Tanimoto. Nos coeficientes de similaridade, o termo “a” representa o número de características presentes em uma molécula A e ausente em uma molécula B, o termo “b” representa o número de características presentes em B, mas ausentes em A, o termo “c” representa o número de características comuns às moléculas A e B. Outro termo aparece nas fórmulas de medidas de similaridade que é o termo “d”, este termo representa o número de características que está ausente em ambas as moléculas.¹¹⁴

Tabela 2. Expressão de alguns coeficientes de similaridade

Coeficiente	Expressão
1. Coeficiente de combinação simples	$(c + d) / (a + b + c + d)$
2. Tanimoto, Jaccard	$c / a + b - c$
3. Dice, Sorensen, Czekanowski, Hodgkin-Richards	$c / 0.5 [(a + c) + (b + c)]$
4. Cosine, Ochiai, Carbo	$c / \sqrt{(a + c)(b + c)}$
5. Russell/Rao	$c / (a + b + c + d)$
6. Rogers/Tanimoto	$(c + d) / (2a + 2b + c + d)$
7. Baroni-Urbani/Buser	$(\sqrt{cd} + c) / (\sqrt{cd} + a + b + c)$
8. Kulczynski-2	$\frac{1}{2} \left(\frac{c}{a + c} + \frac{c}{b + c} \right)$

Para as medidas de similaridade, considerando-se duas moléculas A e B, para cada uma delas uma sequência de “bits” é construída. Cada “bit” indica a presença ou ausência de uma característica estrutural específica (“fingerprint”) da molécula. A característica estrutural específica de uma estrutura química identifica uma molécula

particular através de suas características especiais, da mesma maneira como uma impressão digital identifica um indivíduo.¹¹¹ Assim, através de representações de conjuntos binários (“bits”), duas moléculas podem ser comparadas de acordo com seu grau de similaridade (Figura 5).

Molécula A	1	1	1	0	0	1	4 bits
Molécula B	1	1	0	1	1	1	5 bits
A + B	1	1	0	0	0	1	3 bits

Figura 5. Na representação binária, o nº 1 representa uma característica presente, e o nº 0 uma característica ausente, seja nas estruturas A ou B. O número de “bits” em A indica as características presentes em A e o número de “bits” em B as características presentes em B. Os “bits” em C indicam as características comuns às estruturas A e B

O coeficiente de Tanimoto, por exemplo, fornece valores de comparação na faixa de 0 (zero) a 1. O valor de 0 (zero) significa que não há similaridade entre as moléculas comparadas, já o valor de 1 denota 100% de similaridade, ou seja, como as características moleculares mapeadas representam uma marca única, o valor máximo de 1 garante a comparação de uma molécula a ela mesma ou a seu enantiômero (Figura 6). Através do coeficiente de Tanimoto, calcula-se a razão do número de características comuns entre as moléculas, pelo número total de características presentes em ambas, subtraído pelo número de características comuns (Tabela 2). Um aspecto importante dessa medida é a sua ênfase sobre a presença de características comuns.¹¹⁵

O resultado de uma busca por similaridade é basicamente uma lista de moléculas ranqueadas de acordo com o coeficiente de medida empregado, seguindo uma ordem de similaridade em relação à(s) molécula(s) de referência. Para o coeficiente de Tanimoto, quanto mais próximo seu índice

do valor de 1, mais similar é a molécula avaliada (Figura 6). Normalmente, os valores de similaridade acima de 0,7 são mais representativos, como por exemplo, entre a molécula referência A e os compostos C e D (Figura 6), os quais são derivados da molécula referência e apresentam, basicamente, diferenças entre heteroátomos e substituintes. Índices pouco abaixo de 0,7, normalmente, selecionam moléculas que podem apresentar não somente diferenças em relação aos substituintes, mas também em relação às mudanças no núcleo de heterociclos em relação à molécula referência A, como por exemplo, o composto E (Figura 6). Já índices muito baixos, normalmente, indicam moléculas muito distintas da molécula referência e com poucas características em comum, como por exemplo, o composto F (Figura 6).

Dada a importância atual das técnicas de triagem por similaridade no planejamento de fármacos, há muito interesse no uso e desenvolvimento de medidas de similaridade baseadas em características bidimensionais

estruturais únicas (“fingerprints” 2D), as quais focam na presença ou ausências de fragmentos moleculares nas moléculas avaliadas. Entretanto, vale ressaltar que existem outros tipos de representação estrutural sugeridos para o cálculo de

similaridade intermolecular, entre eles está a representação por propriedades físico-químicas, gráficos químicos, índices topológicos, padrão farmacofórico 3D, dentre outros.¹¹³

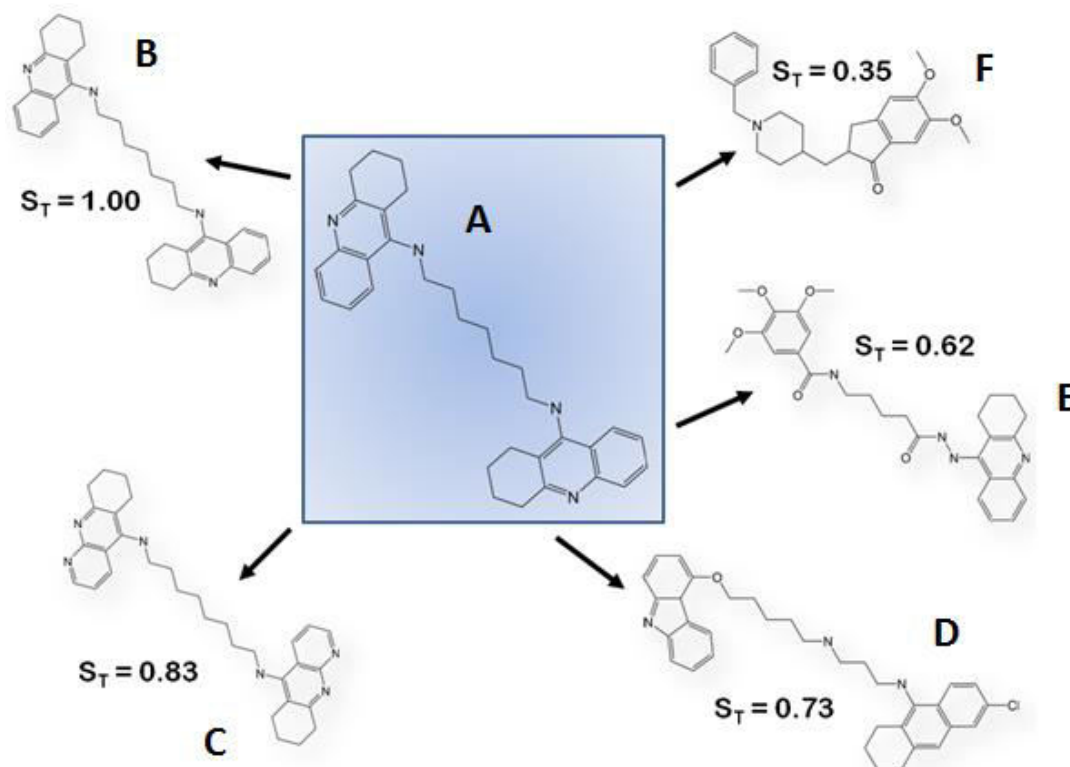


Figura 6. Comparação da similaridade molecular entre a molécula-alvo e outras estruturas através do índice de Tanimoto (S_T)

Embora a triagem por similaridade baseada nestas características únicas (“fingerprint” 2D) seja aplicada com sucesso na busca por compostos bioativos em grandes bases de dados, nem sempre é possível selecionar um método adequado. Duan e colaboradores¹¹⁶, aplicaram 8 métodos distintos destas características únicas bidimensionais em 5 alvos moleculares diferentes empregando um conjunto de dados devidamente validado cujo desempenho frente a diversas classes de moléculas foi avaliado. Cada uma destas características é dotada de propriedades ímpares, que distinguem o seu desempenho. Algumas destas, dizem respeito à presença de fragmentos lineares e fechamento do anel, fragmentos lineares e ramificados,

fragmentos que crescem radialmente a partir de cada átomo, pares de átomos diferenciados por tipo e a distância que os separa, trio de átomos diferenciados por tipo e as três distâncias que os separam ou quatro átomos ligados consecutivamente diferenciados pelo tipo. Com estes estudos, foi possível compreender que a maioria das características bidimensionais estruturais únicas, possui taxas similares de recuperação de moléculas bioativas em bases de dados.

Esta questão levou à consideração do uso de métodos de fusão de dados para combinar os resultados das pesquisas nos bancos de dados que utilizam diferentes coeficientes de similaridade.¹¹⁷ Este método de fusão de dados está sendo cada vez mais utilizado

para melhorar o desempenho de sistemas para triagem virtual em pesquisas farmacêuticas. Existem três tipos de fusão de dados, sendo que o primeiro utiliza a mesma molécula de referência e mesma representação molecular (“fingerprint”) usando diferentes coeficientes de similaridade. O segundo utiliza a mesma molécula de referência e o mesmo coeficiente de similaridade, mas diferentes representações moleculares. O terceiro tipo de fusão, emprega os mesmos “fingerprints” e mesmo coeficiente de similaridade, porém variadas moléculas de referência. Este último é chamado de fusão de grupos. A fusão de grupos é a combinação de uma lista de similaridade feita a partir de vários compostos de referência pertencendo à mesma classe de atividades.

No geral, diferentes medidas de similaridade produzem ranqueamentos diferentes, portanto, os melhores resultados são obtidos utilizando os métodos de fusão de dados para combinar os ranqueamentos resultantes de diferentes coeficientes.¹¹⁸ As triagens por similaridade através da combinação de várias moléculas de referência, normalmente, apresentam resultados superiores à busca convencional, a qual utiliza apenas uma única estrutura bioativa como referência.¹¹⁹⁻¹²¹

A seleção de moléculas baseada na similaridade por forma molecular e também na similaridade por distribuição eletrostática desempenha um importante papel no reconhecimento molecular do sítio receptor e contribui de maneira considerável para a seleção de ligantes com afinidade pelo alvo molecular.¹²¹ Muitas representações moleculares e medidas de similaridade estão disponíveis para auxiliar no desenvolvimento de fármacos e a qualidade da série de compostos resultantes é altamente dependente da forma tridimensional e da distribuição eletrostática da molécula (eletroforma)¹²².

Em um estudo para a seleção de moléculas baseado na similaridade de forma e na similaridade eletrostática, Tresadern e

colaboradores¹²³ fizeram uma comparação de métodos de triagem virtual baseado no ligante e suas aplicações para o receptor do fator 1 de liberação da corticotrofina (do inglês Corticotrofin Releasing Factor 1 Receptor, CRF1)¹²⁰. Empregando em seus estudos, programas que adotam o uso de fingerprints 3D como o programa ROCS (do inglês, Rapid Overlay of Chemical Structures) que usa o coeficiente de Tanimoto para calcular similaridade por forma,¹²⁴ e outro grupo, com o programa OpenEye Combo-Score e o programa EON que usam o coeficiente de Tanimoto para cálculo de similaridade eletrostática.¹²⁵, obtiveram um bom desempenho em identificar os ligantes ativos, e os métodos 3D se destacaram por ter obtido um melhor desempenho frente aos métodos 2D.

A similaridade molecular é um dos temas mais discutidos em química medicinal, pois existem muitas medidas de similaridade que estão disponíveis para auxiliar o pesquisador na descoberta de novos fármacos através da aplicação de técnicas de triagem virtual, assim como na criação de um novo conjunto de dados de compostos e pela disponibilidade de diversos coeficientes de medida de similaridade descritos na literatura.

A qualidade da base de dados ou dos compostos selecionados está diretamente ligada e é dependente da representação molecular e da medida de similaridade que será selecionada para o estudo. Existe uma variedade de maneiras de se medir a semelhança entre as moléculas e diferentes tipos de coeficientes de similaridade. Nos últimos anos houve um rápido desenvolvimento de muitas abordagens computacionais no desenvolvimento de medidas de similaridade.

Entretanto, não há um consenso sobre qual método seja o melhor, pois ainda se faz necessário a validação de tais procedimentos e o uso de sistemas de filtragem para remover as moléculas com características indesejáveis. Sendo assim, uma abordagem racional na escolha dos coeficientes a serem

utilizados para um estudo de similaridade ou de triagem virtual baseada em similaridade deve basear-se em técnicas de fusão de dados, onde há uma combinação dos resultados das pesquisas nos bancos de dados que utilizam diferentes coeficientes de similaridade, além de ser levado em consideração um estudo sobre a cavidade de ligação do alvo biológico na qual o ligante terá que se ajustar, pois existe uma relação de dependência entre a similaridade e a macromolécula biológica onde os compostos análogos se ligam.

4.3. Descritores e novas tendências

Um descritor molecular é uma representação matemática de uma molécula, resultante de um procedimento lógico no qual a informação estrutural é codificada e convertida em uma representação simbólica. O conteúdo da informação de um descritor molecular depende da representação molecular do composto (Tabela 3) e do algoritmo que é empregado para o cálculo do descritor.¹¹⁴ Estratégias que empregam descritores constituem as novas tendências no processo de triagem virtual de substâncias bioativas, incluindo as técnicas que envolvem a aprendizagem de máquina (Machine Learning).¹²⁶⁻¹²⁸

Tabela 3. Classificação dos descritores de acordo com a representação molecular¹¹⁴

Representação Molecular	Descritores
0D	Peso molecular, número de átomos, número de ligações, soma de propriedades atômicas.
1D	Número de fragmentos.
2D Descritores topológicos	Índice de Zagreb, Índice de Wiener, descritores BCUT, vetor de autocorrelação 2D.
3D Descritores geométricos	Descritor 3D-MORSE, descritores WHIM, descritores GETAWAY, vetores de autocorrelação 3D.
3D Propriedades da superfície	Potencial eletrostático, potencial de hidrofobicidade, potencial de ligação de hidrogênio.
3D Propriedades da grade	Análise comparativa dos campos moleculares (CoMFA).
4D	Coordenadas 3D e amostragem de conformações, valores da energia de interação, descritores GRIND.

A área de atuação de descritores moleculares é em grande parte interdisciplinar. Para o uso destas técnicas é necessário o conhecimento de abordagens estatísticas, de quimiometria (uma área que se refere a problemas de quantificação envolvidos com regressão multivariada) e dos princípios de QSAR (relação quantitativa

entre estrutura e atividade biológica, do inglês “quantitative structure-activity relationship”), e de QSPR (relação quantitativa entre estrutura e propriedade, também do inglês “quantitative structure-property relationship”). Os métodos estatísticos, como por exemplo, a análise de regressão multilinear e redes neurais

artificiais, são métodos utilizados na construção de modelos. Esses modelos relacionam os descritores com a propriedade ou a atividade biológica dos compostos de interesse e podem ser aplicados em triagem virtual indicando a potencial atividade das estruturas de uma base de dados. Para isso, é desejável que o conjunto de estruturas a partir do qual o modelo QSAR e/ou QSPR é obtido (conjunto de treinamento) apresente uma diversidade estrutural representativa.¹²

Um dos principais desafios é desenvolver compostos com as propriedades químicas, físicas e biológicas desejadas. Uma grande variedade dessas propriedades é totalmente dependente da estrutura tridimensional da molécula, portanto os estudos de QSAR/QSPR estão ganhando cada vez mais destaque. Abordagens que envolvem a modelagem e a predição da atividade biológica, assim como a triagem virtual e experimentos de docagem ou “docking” (predição da interação receptor/ligante e complexos em sistemas biológicos) necessitam da informação da estrutura tridimensional das moléculas. Entretanto, as propriedades químicas, físicas e biológicas de um composto orgânico, normalmente, não podem ser preditas simplesmente de maneira direta através do conhecimento de sua estrutura molecular, sendo necessário o estabelecimento de abordagens indiretas de QSAR e QSPR. Tais estratégias são indutivas, no sentido de que um modelo é gerado através do conhecimento das propriedades ou atividades biológicas de uma série de compostos químicos. Em geral, os estudos de QSPR/QSAR começam a partir de um conjunto de moléculas bioativas e os descritores calculados devem, no mínimo, extrair informações sobre a topologia molecular de cada composto.¹¹⁴

A grande variedade de descritores topológicos (2D) existentes são aplicados frequentemente em modelagem física, química ou de propriedades biológicas de compostos orgânicos. Estes descritores não dependem de orientação ou conformação, mas sim de características moleculares que

refletem apenas a composição em caráter bidimensional da molécula (Tabela 3).¹²⁹

Se os descritores moleculares são derivados de representações moleculares 3D, esses perfis podem ser obtidos experimentalmente ou, quando não disponíveis, calculados por métodos computacionais. Assim, cresce a importância do conhecimento das conformações bioativas das moléculas do conjunto de treinamento. As geometrias calculadas por métodos computacionais para as moléculas das bibliotecas virtuais são submetidas a uma análise conformacional a fim de restringir o estudo às conformações mais viáveis, ou seja, de baixa energia. Assim, os descritores calculados devem refletir parâmetros inerentes à conformação das moléculas relacionados à estrutura tridimensional.¹¹⁴

A análise comparativa dos campos moleculares (CoMFA, do inglês “Comparative Molecular Field Analysis”), um dos métodos de cálculo de QSAR-3D, desenvolvido por Cramer e colaboradores,¹³⁰ descreve a relação entre as características moleculares estéricas e eletrostáticas 3D de uma série de ligantes com afinidades de ligação conhecidas, com o mesmo mecanismo de ação, na mesma cavidade de ligação de um receptor. Esta análise é realizada por meio de um algoritmo de sobreposição flexível, que realiza a sobreposição de ligantes com o objetivo de predizer afinidades de ligação para novos compostos que sejam similares ao grupo de moléculas de referência.

Em uma grade retangular (“grid”) as medidas de energias de ligação são efetuadas com átomos de prova colocados consecutivamente em cada ponto da grade. Assim, as afinidades de ligação experimentais são correlacionadas com a contribuição dos campos moleculares medidos na grade usando um algoritmo de aprendizagem estatística (análise dos mínimos quadrados parciais - PLS, do inglês “Partial Least Squares”), produzindo uma equação de regressão.¹³¹

Uma abordagem alternativa ao CoMFA

são os métodos de análise comparativa de índices de similaridade molecular (CoMSIA, do inglês “Comparative Molecular Similarity Indices Analysis”) que fazem o cálculo dos campos de propriedades expressos em índices de similaridade molecular. Os campos de diferentes propriedades físico-químicas no CoMSIA usam índices de similaridade que dependem da distância e não ocorrem singularidades nas posições atômicas. No CoMFA, são necessários valores limiares nos campos e os mapas de contribuição dos campos gerados, muitas vezes, não são ligados de forma contígua, isso faz com que a sua interpretação torne-se difícil. Os mapas obtidos nos métodos de CoMSIA são superiores e fáceis de serem interpretados e sua principal característica é destacar as regiões dentro da área ocupada pelos ligantes que necessitam de uma determinada propriedade físico-química que será importante para a atividade biológica, ao invés de designar regiões separadas das moléculas no qual presumia-se que ocorreriam as interações, como ocorre nos métodos de CoMFA.¹³²

Em um trabalho desenvolvido por Thondorf, I. e colaboradores (2011), foi realizado um estudo de QSAR-3D em substratos sintéticos e naturais do transportador de aminoácidos acoplados a prótons (hPAT1). A análise comparativa de índices de similaridade molecular (CoMSIA) foi aplicada para verificar cuidadosamente os requisitos da estrutura para a ligação com os substratos e também para derivar um modelo de previsão que pudesse ser utilizado para a concepção de novos pró-fármacos. Os estudos de QSAR-3D empregados neste trabalho foram utilizados como uma ferramenta para elucidar o mecanismo de ação do fármaco a nível molecular uma vez que a estrutura tridimensional está relacionada com a afinidade ao receptor. Os dados de 50 substratos de hPAT1 foram usados para os estudos de CoMSIA na derivação de um modelo matemático de relação estrutura-atividade e para identificar quais as características estruturais que determinam a ligação desses substratos à proteína. A conformação biologicamente

ativa dos compostos foi alinhada de maneira a representar um modo de ligação semelhante. Em análises de QSAR-3D o alinhamento das moléculas é uma etapa muito importante, este alinhamento é feito pela sobreposição de características estruturais comuns usando o confôrmero de menor energia de cada estrutura ou a conformação bioativa encontrada na estrutura cristalográfica resolvida por raios-X quando esta estiver disponível¹³³. Os modelos de CoMSIA foram gerados usando cinco descritores (estérico, eletrostático, hidrofóbico, doador e acceptor de ligação de hidrogênio) como variáveis independentes e os valores de pK_i como variáveis dependentes realizando diferentes combinações. Subsequentemente foi realizada a análise de regressão dos mínimos quadrados parciais (PLS, do inglês “Partial Least Square”). Assim, os parâmetros estatísticos do CoMSIA juntamente com a contribuição dos descritores variáveis foram resumidos em cinco modelos que foram classificados como sendo os melhores, caracterizados por significativas propriedades correlativas e preditas. A triagem virtual foi realizada em uma base de dados de aproximadamente 650.000 moléculas utilizando um modelo de farmacóforo derivado dos mapas de contorno do CoMSIA permitindo, assim, a identificação de possíveis substratos de hPAT1. As afinidades dos compostos foram preditas e 11 compostos foram identificados como possíveis substratos de alta afinidade. Dos 11 compostos, dois compostos selecionados inibiram fortemente a proteína. O modelo de QSAR-3D empregado neste trabalho para racionalizar e entender as afinidades de ligação dos substratos de hPAT1 fornece uma introspecção das interações entre os substratos e o seu alvo biológico (hPAT1) permitindo, assim, a predição da afinidade desses novos compostos¹³³.

Desenvolvido por Hopfinger,^{134,135} o QSAR-4D é um método de modelagem molecular que é muito útil na construção de modelos farmacofóricos quantitativos 3D para um conjunto de ligantes, quando a geometria do receptor correspondente não é conhecida.

No QSAR-4D, uma grade é utilizada ao redor das estruturas dos ligantes para determinar as regiões no espaço tridimensional que são responsáveis por possíveis interações, mas *a priori* nenhuma prova ou energia de interação é utilizada. Em vez disso, este método incorpora flexibilidade molecular e alinhamentos múltiplos, permitindo, assim, a identificação da conformação que potencializa a atividade biológica prevista.

Os métodos de QSAR multidimensional, os quais incluem os métodos de QSAR-4D, 5D, 6D e 7D são essencialmente extensões dos métodos de QSAR-3D. Estes métodos incorporam características físicas ou propriedades adicionais necessárias para contornar dificuldades provenientes dos métodos de QSAR-3D.¹³⁶

Através do QSAR multidimensional a atividade dos compostos pode ser relacionada ao descritor molecular em amostras de representações moleculares tridimensionais baseadas em dados atomísticos modelados *in silico* e/ou medidos por um método experimental (como por exemplo, em análises de raios-X). Recentemente, tem sido demonstrado que o QSAR-5D e o 6D podem ser usados para representações múltiplas do receptor, bem como de seus estados de solvatação. No método de QSAR-5D, Vedani e colaboradores¹³⁷ e Lill e colaboradores¹³⁸ introduziram uma representação múltipla da hipótese de ajuste induzido, onde a adaptação do sítio de ligação do receptor à topologia individual do ligante era considerada como uma quinta dimensão.

Estas dimensões de que trata o QSAR multidimensional referem-se ao tipo de informação codificada pelo descritor, que é usada como uma variável independente. Vedani e Lill propuseram a seguinte classificação: método de QSAR-1D, no qual as propriedades globais dos ligantes (exemplo, pKa, logP) são correlacionadas com a afinidade dos compostos; método de QSAR-2D, o qual usa “fingerprints” estruturais do ligante, por exemplo, conectividade química, negligenciando a informação estrutural

tridimensional; QSAR-3D, a representação tridimensional da estrutura é investigada para provar os descritores moleculares; QSAR-4D, devem considerar orientações/conformações de ligantes múltiplos; QSAR-5D, consideram adicionalmente representações múltiplas do efeito de ajuste induzido do receptor/ligante; QSAR-6D, considera adicionalmente, modelos múltiplos de solvatação. Formalmente, pode-se adicionar a essa classificação o modelo de QSAR-0D, em que um único parâmetro é correlacionado com a afinidade dos compostos. Dessa forma, a dimensão de QSAR (Tabela 4) refere-se à representação estática do ligante em QSAR-3D, representações múltiplas do ligante em QSAR-4D, modelo de receptor ou pseudo receptor virtual baseado no ligante em QSAR-5D, cenários de múltipla solvatação em QSAR-6D e dados do receptor real ou interação ligante-receptor no QSAR-7D. O protocolo de QSAR-7D deve também incluir modelos virtuais de receptor baseado no alvo, obtidos por modelagem de proteínas homólogas.

Deve-se observar que quanto mais complexo o descritor molecular considerado, maior também é a dimensão do QSAR, portanto, esta dimensão é um número artificial que não pode ser relacionado diretamente à topologia molecular. Em vez disso, este sistema tem como objetivo classificar uma variedade de métodos que são incluídos em QSAR.¹³⁹

4.4. Aprendizagem de máquinas

A Aprendizagem de máquinas (“Machine Learning Approach”) é um ramo da inteligência artificial que estuda o desenvolvimento de algoritmos capazes de classificar objetos a partir de um dado conjunto de treinamento. Em quimioinformática, os objetos normalmente são moléculas e as classes são propriedades inerentes a estas. Um exemplo é o uso de aprendizagem de máquinas para discriminar

uma série de moléculas em ativas ou inativas em relação a um determinado alvo molecular. Estas técnicas podem prever padrões numéricos a partir de um conjunto

de treinamento, como por exemplo, valores de atividade inibitória (IC_{50}), que fornece a estrutura dos compostos ativos e seus respectivos valores de IC_{50} .¹⁴⁰

Tabela 4. Classificação geral do QSAR multidimensional

Dimensão de QSAR	Representação molecular usada para cálculos de descritor molecular
0D	Parâmetro de número único geralmente referindo-se a propriedade global.
1D	Uma combinação multivariada de propriedades globais.
2D	Padrões estruturais relativos às representações moleculares 2D.
3D	Único estado de representação 3D de ligantes.
4D	Múltiplas representações 3D de ligantes.
5D	QSAR-3D + cenários de ajuste induzido.
6D	QSAR-5D + cenários de múltipla solvatação.
7D	Receptor real ou dados do modelo do receptor baseado no alvo.

Para a obtenção de bons resultados utilizando a técnica de aprendizagem de máquinas é de suma importância a escolha de um conjunto de treinamento adequado, que represente de maneira consistente as propriedades a serem classificadas. O conjunto de treinamento, é constituído por um grupo finito de exemplos distribuídos dentro de um sistema com a finalidade de se encontrar uma classificação padrão desejada (propriedade em estudo). Este conjunto de treinamento é aplicado ao algoritmo de aprendizagem escolhido, a fim de treinar este algoritmo para que seja capaz de classificar novos dados dentro deste sistema. Uma vez treinado o algoritmo, um conjunto teste (conjunto de exemplos não incluído no conjunto de treinamento) é utilizado para validar o método.¹⁴⁰

Atualmente, há grande interesse na aplicação dessas ferramentas e técnicas para obtenção de respostas em modelos biológicos e farmacológicos^{140,141} incluindo modelos de QSAR e QSPR,^{5,141-145} bem como simulações de docagem e triagem virtual.^{121,146-148}

Em triagem virtual, a aprendizagem de máquinas tem sido empregada com resultados semelhantes ou, em alguns casos, até mesmo melhores do que os métodos convencionais utilizados, com um custo computacional reduzido.¹⁴⁹ Os algoritmos de aprendizagem de máquinas aplicados em triagem virtual são concentrados no aprendizado supervisionado, onde um conjunto de treinamento é fornecido para o aprendizado do sistema e os objetos devem ser classificados como pertencentes a este conjunto.

Dentre os algoritmos de aprendizagem de máquinas, os mais empregados em triagem virtual são: as redes neurais artificiais (ANN, do inglês “artificial neural network”), máquinas de vetores suporte (SVM, do inglês “support vector machine”) e o k -ésimo vizinho mais próximo (k -NN, do inglês “ k -nearest neighbors”).

Técnicas de aprendizagem de máquinas têm sido cada vez mais aplicadas em triagem virtual baseada em ligantes, devido à expansão das bibliotecas químicas, do surgimento de novos descritores moleculares

e das técnicas de busca por similaridade. Além disso, com a maior facilidade de obtenção de dados de atividade em ensaios biológicos, os mesmos vêm sendo incorporados em modelos que são usados para auxiliar no planejamento de novos candidatos a fármacos, o que facilita a identificação de características moleculares que são relevantes para a propriedade biológica, direcionando, assim, a síntese de compostos promissores.¹⁴⁹

4.4.1. Redes neurais

Uma rede neural artificial (ANN, do inglês “artificial neural network”) é um modelo matemático inspirado nas redes neurais biológicas. As ANN são unidades que processam informações de uma forma motivada pela funcionalidade do sistema nervoso biológico. Assim, como o cérebro consiste de neurônios que estão ligados uns aos outros, uma ANN abrange neurônios artificiais interconectados. Um neurônio artificial é construído em analogia aos neurônios biológicos. Os neurônios trabalham em conjunto para resolver um determinado problema. Assim, como em humanos, as ANN aprendem com os exemplos (dados de entrada). O processo de aprendizagem de uma ANN é chamado treinamento. No cérebro humano, as sinapses (conexões entre os neurônios), são adaptadas durante um processo de aprendizagem. Um processo correspondente é executado quando uma ANN é treinada, onde os pesos de suas conexões são ajustados com o intuito de reproduzir os exemplos disponibilizados como dados de entrada. Depois do aprendizado, elementos a serem preditos são adicionados com o intuito de estimar sua possível atividade biológica.¹¹⁴

As redes neurais artificiais têm diversas aplicações no processo de planejamento de substâncias bioativas, podendo ser empregadas, inclusive, para a determinação da relação quantitativa entre estrutura molecular e atividade biológica (QSAR) de

candidatos a fármacos.¹²⁶

As ANN já foram utilizadas com relativo sucesso na predição de inibidores de HIV-protease,¹⁵⁰ combinando um algoritmo genético para a seleção de variáveis usadas para estimar a afinidade de derivados de arilpiperazinas ao receptor 5-HT1A, apresentando uma taxa de erro percentual baixa, cerca de 0,065.¹⁵¹ Sucesso também pôde ser observado na predição da permeabilidade à barreira hematoencefálica e ligação à soroalbumina, onde os valores preditos apresentaram um coeficiente de determinação (R^2) na ordem de 0,9 frente a 126 fármacos comerciais.¹⁵² Além disso, ANNs já foram empregadas na predição de inibidores da Furina, capazes de evitar a maturação das toxinas produzidas pelo *Bacillus anthracis*,¹⁵³ e, também, aplicadas na geração de um modelo de QSAR de predição de toxicidade de pirril-aril-sulfonas, utilizadas como inibidores não nucleosídicos de transcriptase reversa no tratamento da AIDS, com uma taxa de erro de 0,066 frente às concentrações citotóxicas obtidas e as calculadas.¹⁵⁴

Os mapas auto-organizáveis (do inglês, “Self-Organizing Maps” - SOM) constituem um tipo de ANN que transforma um espaço multidimensional em bidimensional, classificando os dados devido a sua similaridade, já que os mesmos são agrupados de acordo com as suas características. Este método foi usado com sucesso por Fernandes e colaboradores¹⁵⁵ para classificar lactonas sesquiterpênicas quanto à sua atividade contra carcinoma nasofaríngeo, assim como predizer a toxicidade destes compostos, atingindo um índice de 89% de poder preditivo.

Os SOM também já foram testados frente a bibliotecas de inibidores de ciclo-oxigenase 2 (COX-2), acetilcolinesterase, fosfodiesterase IV (PDE4), trombina 2105 e inibidores de ativação de u-plasminogênio (uPA) 199, e apresentaram sucesso de aproximadamente 70% frente aos compostos com atividade conhecida.¹⁵⁶ Além disso, observa-se o emprego desse método na

identificação de inibidores de glicoproteína-P, elucidando quinazolinonas, benzotiazóis e propafenonas com atividade experimental na faixa de concentração de μM .¹⁵⁷

4.4.2. "Support vector machine"

O SVM ("Support Vector Machine")¹⁵⁸ é um algoritmo baseado na teoria de aprendizado estatístico supervisionado que consiste em uma técnica computacional de aprendizado para problemas de reconhecimento de padrões.¹⁵⁹ Ela utiliza a dedução de dados a partir do conjunto de treinamento, e tem a habilidade de classificar os objetos de um determinado espaço altamente dimensional em duas classes distintas, traçando um hiperplano, de acordo com suas características, de tal forma que a separação entre os exemplos seja máxima.^{160,161}

SVM foi originalmente desenvolvido para classificação binária, a qual busca a construção de um hiperplano como superfície de decisão. Dentro desta técnica

podemos encontrar dois tipos de conjunto de informações: os que apresentam um padrão linearmente separável e os que apresentam um padrão não-linearmente separável. No segundo tipo, busca-se uma função de mapeamento apropriada para tornar o conjunto mapeado linearmente separável.¹⁶²

O termo vetor é definido como o conjunto de características que descrevem um caso, ou seja, uma linha de valores de predição. Assim, o objetivo da modelagem SVM é encontrar o hiperplano ótimo que separa grupos de vetores de tal forma que os casos com uma categoria de alvo variável estão em um lado do plano e os casos com a outra categoria estão no outro lado do plano (Figura 7). Existem infinitas retas capazes de separar os dois conjuntos de treinamento. As retas mostradas em verde (Figura 7) representam os possíveis hiperplanos, sendo que a reta mais espessa é a que mais bem representa o melhor posicionamento para separação dos grupos, sendo esta chamada de hiperplano ótimo. As retas paralelas a este hiperplano são as que definem as margens e os vetores de suporte.

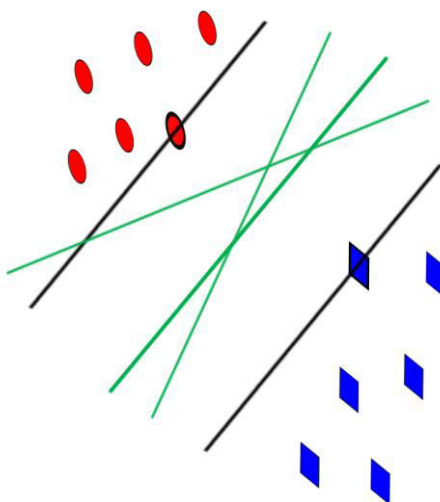


Figura 7. Representação geral do processo SVM para determinação do hiperplano ótimo. Adaptado de Gonçalves (2010)¹⁶¹

Se X_i ($i=1, 2, \dots, M$) é um conjunto de treinamento em um problema, os quais estão presente em dois grupos linearmente separáveis (ω_1 e ω_2), a cada objeto fica

associado um rótulo: $y_i=1$ se $X_i \in \omega_1$, $y_i=-1$ se $X_i \in \omega_2$. Nesse caso a função de decisão linear (hiperplano ótimo) adquire a forma:

$$D(x) = w \cdot x + b,$$

onde w é um vetor m -dimensional (pesos) e b é o termo independente (*bias*), para $i=1, 2, \dots, M$.¹⁶³

Dentro da linguagem de SVM, podemos definir a margem como a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado para separação destes grupos.¹⁶⁴ A margem determina o quanto dois grupos podem ser

separados,¹⁶² e os vetores que determinam estas margens são denominados de vetores suporte (Figura 8). Assim, temos que as funções que definem as margens são:

$$\text{margem 1: } D(x) = w \cdot x_i + b = 1$$

$$\text{margem 2: } D(x) = w \cdot x_i + b = -1$$

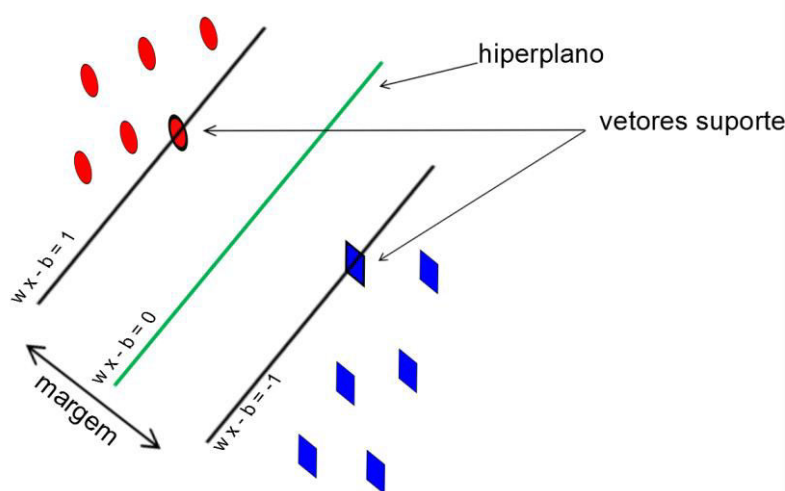


Figura 8. Identificação da margem e dos vetores suporte sobre a linha pontilhada. Adaptado de Lima (2002)¹⁶⁵

Uma vez determinados os vetores suporte, estes se tornam valores de referência, que sozinhos determinam o hiperplano ótimo, sendo que os demais vetores se tornam irrelevantes, podendo ser removidos do conjunto de treinamento sem afetar os resultados.¹⁶² Em problemas reais, os padrões linearmente separáveis são dificilmente encontrados, sendo a maioria deles complexos e não-lineares. Para estender a SVM linear a resolução de

problemas não-lineares foram introduzidas funções reais, que mapeiam o conjunto de treinamento em um espaço linearmente separável, o espaço de características. Um conjunto de dados é dito ser não-linearmente separável caso não seja possível separar os dados com um hiperplano definido por uma função do tipo $f(x) = w \cdot x + b$. A figura 9 mostra um conjunto linearmente e outro não-linearmente separável.

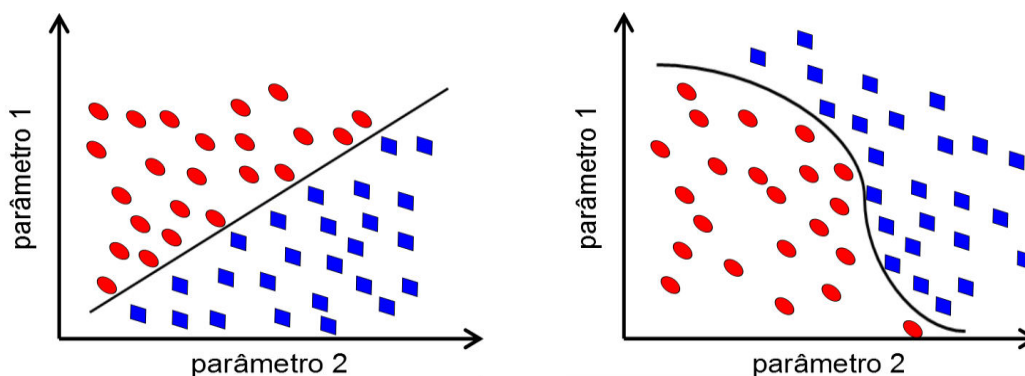


Figura 9. Exemplos de padrões linearmente e não-linearmente separáveis, respectivamente. Adaptado de Zien e colaboradores¹⁶⁶

O teorema de Cover afirma que um problema não-linear tem maior probabilidade de ser linearmente separável em um espaço de mais alta dimensionalidade.¹⁶² A partir disso, a SVM não-linear realiza uma mudança de

dimensionalidade, por meio da função de Kernel (Φ), caindo então em um problema de classificação linear (Figura 10), podendo fazer uso do hiperplano ótimo que agora é definido por $f(x) = (w \cdot x) \cdot \Phi + b$.¹⁶⁵

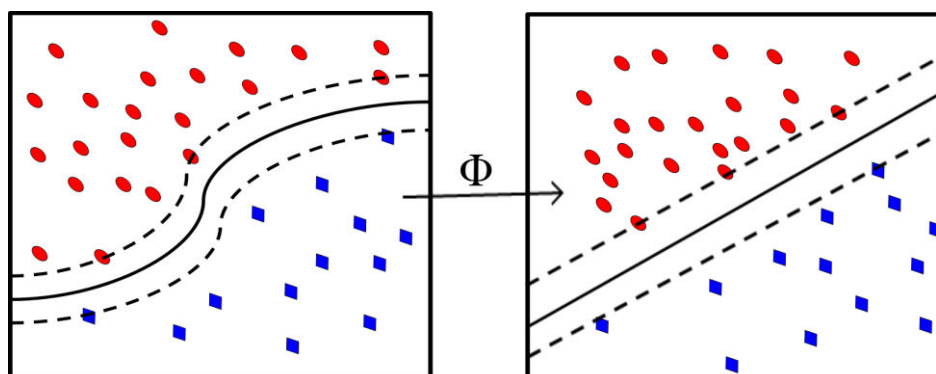


Figura 10. Representação esquemática da mudança de um padrão não - linearmente separável para um linearmente separável através da aplicação da função Kernel (Φ). Adaptado de Zien e colaboradores¹⁶⁶

Devido a sua eficiência em trabalhar com dados de alta dimensionalidade, SVM é relatada na literatura como uma técnica altamente robusta, muitas vezes comparada às Redes Neurais.^{167,168} SVM já foi usado para a predição de inibidores de HIV-protease,¹⁵⁰ com melhores resultados que os obtido com as ANNs ("artificial neural network"). Han e colaboradores¹²⁷ realizaram um estudo com o intuito de verificar a aplicação da SVM em triagem virtual com inibidores de HIV protease, diidrofolato redutase (DHFR) e

antagonistas de dopamina, onde dois conjuntos de treinamento foram fornecidos ao SVM: um com compostos ativos e outro com compostos inativos, sendo que esses compostos inativos tinham estruturas semelhantes aos ativos. Os resultados obtidos pelo SVM quando comparados a outros métodos de triagem baseados em ligantes mostrou um índice de eficiência semelhante. A etapa determinante para o bom resultado da SVM, sem dúvida alguma, é a seleção do conjunto de treinamento, pois

este pode afetar sobremaneira o resultado final.

Sucesso também pôde ser observado na predição de inibidores seletivos para a enzima conversora do fator de necrose tumoral (TACE),¹⁶⁹ apresentando alta eficiência, cerca de 99% de precisão para determinar os compostos inibidores e 98% para determinar compostos não-inibidores, desde que os descritores fossem bem selecionados. Além disso, observa-se também o emprego do SVM de forma combinada ao k-NN na classificação de inibidores e não inibidores de CYP1A2 (uma isoforma do citocromo P450), alcançando níveis de precisão de 73% a 76% sobre o conjunto de teste.¹⁷⁰

Liew e colaboradores¹⁷¹ propuseram criar um modelo de SVM para realização de triagem virtual de inibidores de Lck (Lymphocyte-specific protein tyrosine kinase). Para tal, foram utilizados 66.032 compostos de 8.423 famílias químicas, reunidas a partir de estudos publicados na literatura e das bases de dados de estruturas e atividades biológicas do PUBCHEM e do MDDR. Inibidores desta enzima têm potencial emprego no tratamento de doenças auto-imunes e rejeições a transplantes. Um total de 100 descritores moleculares foi listado nessa investigação, sendo 13 referentes a propriedades moleculares simples, 13 descritores de carga, 34 de conectividades moleculares e descritores de forma e 40 índices de estado eletrotológico. O modelo de SVM se mostrou capaz de identificar novos inibidores de Lck de uma grande biblioteca de compostos químicos. Esta biblioteca de inibidores possuía um índice de 0,52% de falsos positivos e foi desenvolvido a partir de um grande conjunto de inibidores e não inibidores de Lck. O uso do SVM, que não depende apenas das associações de famílias químicas, permitiu distinguir não-inibidores que são estruturalmente similares aos inibidores já conhecidos. Muitos outros trabalhos vêm sendo desenvolvidos utilizando a técnica de SVM em triagem virtual.^{149,172-179} Em tais estudos vêm se

comprovando a alta eficiência desta técnica e a sua grande vantagem em relação ao seu poder preditivo quando comparada aos métodos mais clássicos de triagem virtual.

4.4.3. "k-Nearest Neighbors (k-NN)"

O algoritmo k-NN é um método para classificação de objetos baseado em exemplos, também chamado de "aprendizado preguiçoso", onde a função só é aproximada localmente e toda a computação é protelada até a classificação. Ele está dentre os mais simples de todos os algoritmos que envolvem aprendizagem de máquinas: um objeto é classificado por uma maioria de votos dos seus vizinhos, com o objeto sendo designado à classe mais comum entre seus k mais próximos vizinhos (k é um valor positivo, geralmente variando de 1 a 10).¹⁴⁰ Se k = 1, então o objeto é designado à classe de seu único vizinho mais próximo, se k = 2, o objeto é designado à classe de seus dois vizinhos mais próximos e assim por diante. Os vizinhos são obtidos de um conjunto de objetos para os quais a classificação correta é conhecida. Isto pode ser considerado o conjunto de treinamento do algoritmo, embora não seja necessária nenhuma etapa explícita de treinamento.¹⁸⁰⁻¹⁸²

A ideia desse algoritmo é a de que a classe a qual um determinado composto pertence pode ser definida pela classe à qual pertencem os vizinhos mais próximos (compostos mais similares, por exemplo), considerando similaridades ponderadas entre o composto e seus vizinhos mais próximos.¹²⁸

Com o aumento do uso do k-NN, algumas mudanças foram realizadas com o intuito de aprimorar este método, sendo que a mais comum delas é a de considerar a distância entre a molécula em estudo e os seus vizinhos, tendo maior peso as atividades dos vizinhos mais próximos.¹⁴⁰ Para isso foi necessária a aplicação de uma função de ponderação (normalmente uma função

"kernel"). Jensen e colaboradores¹⁸³ utilizaram uma função de ponderação gaussiana da distância de Tanimoto entre as moléculas para prever a atividade de 1,000 inibidores do citocromo P450 2D6 e 3A4. Cedeño e Agrafiotis¹⁸⁴ empregaram uma função kernel do inverso da distância (com um aditivo de correção para evitar singularidades) em um estudo QSAR.

O maior refinamento realizado para este método é a consideração do peso da contribuição de cada descritor para o cálculo da distância. Miller e colaboradores¹⁸⁵ descreveram uma variação na classificação k-NN, métrica flexível modificada do vizinho mais próximo (do inglês, "modified flexible metric nearest neighbor" - MFMNN), onde os pesos dos descritores são determinados pelo quanto o descritor de cada um dos vizinhos está distante do objeto em estudo, sendo que os descritores dos vizinhos mais próximos apresentam maior peso.

Shen e colaboradores¹⁸⁶ demonstraram a eficácia desse algoritmo em vários conjuntos de dados, incluindo seu uso para identificar nove potenciais anticonvulsivantes de uma biblioteca virtual contendo mais de 250 mil moléculas, das quais sete foram confirmadas como bioativas. O mesmo grupo também usa o k-NN como o algoritmo de aprendizagem de seus métodos COLIBRI¹⁸⁷ e ENTess¹⁸⁸ os quais incluem descritores que representam o receptor e, também, o ligante.

O k-NN tem sido amplamente empregado nos últimos anos em vários grupos de pesquisa com diferentes finalidades, como por exemplo: Du-Cuny e colaboradores¹⁸¹ verificaram a influência de modificações químicas sobre a permeabilidade celular de possíveis inibidores de AKT (proteína cinase B); Peterson e colaboradores¹⁸² em estudos de QSAR e triagem virtual, encontraram inibidores de geranyl transferase tipo I, testados com sucesso posteriormente *in vitro* sendo que sete deles com potência na escala milimolar; Vasanthanathan e colaboradores¹⁷⁰ empregaram o k-NN de forma combinada ao SVM na classificação de inibidores e não inibidores de CYP1A2, conseguindo de 73% a 76% de precisão;

Wang e colaboradores¹⁸⁹ combinaram o k-NN a diferentes métodos de cálculo de descritores em estudos de QSAR, para determinação de afinidade e seletividade de agonistas e antagonistas de receptores 5HT1E e 5HT1F, a partir de uma triagem em uma biblioteca virtual de 65 mil compostos, obtendo uma correlação entre pKi teórico e experimental de até 92%; Hsieh e colaboradores¹²⁸ utilizaram o algoritmo na classificação de ligantes e não ligantes de β -lactamases do tipo AmpC, com eficiência de 90%, encontrando 5 compostos com atividade inibitória na escala milimolar. Existem, ainda, diversos outros exemplos do emprego do algoritmo k-NN com sucesso no planejamento de substâncias bioativas.^{150,169,180,190}

5. Conclusões

As técnicas de triagem virtual representaram um grande avanço no processo de planejamento de fármacos, uma vez que tornou mais dinâmico o processo de busca por substâncias bioativas de interesse farmacêutico. As técnicas existentes atualmente têm se mostrado muito úteis, e a cada dia, o número de publicações relatando o sucesso na busca por substâncias bioativas utilizando a triagem virtual vêm aumentando. Entretanto, a cada dia também cresce o número de moléculas sintetizadas e de novos alvos biológicos atrativos descobertos, o que acarreta a necessidade de um processo de triagem de moléculas bioativas mais eficiente, seja ele virtual ou não. Neste contexto, torna-se necessário o desenvolvimento de novas estratégias que consigam traduzir cada vez mais os dados químicos de substâncias de interesse em informações capazes de nortear os cientistas na busca de novos candidatos a fármacos. O desenvolvimento das técnicas de triagem é, à primeira vista, promissor e a cada dia novas estratégias são descritas na literatura, sobretudo desenvolvimentos em relação aos métodos que envolvem aprendizagem de máquinas e análise de descritores, que

trouxeram novas perspectivas para a abordagem da triagem virtual, principalmente por aperfeiçoar os estudos de relação estrutura-atividade e introduzir métodos de análise capazes de detectar semelhanças entre compostos químicos que vão além da simples organização estrutural dos átomos no espaço tridimensional. Existe grande expectativa de que o contínuo aprimoramento destes processos ajudará a reduzir custos e o tempo necessários para o desenvolvimento de novos fármacos, sobretudo para doenças que ainda não têm medicamentos efetivos para seu tratamento, como por exemplo, a doença de Alzheimer, a doença de Chagas e vários tipos de câncer.

Referências Bibliográficas

- ¹ Buchanan, S. G. *Curr. Opin. Drug. Discov. Devel.* **2002**, *5*, 367. [PubMed]
- ² Lima, L. M. *Quim. Nova* **2007**, *30*, 1456. [CrossRef]
- ³ Verli, H.; Barreiro, E. J. *Quim. Nova* **2005**, *28*, 95. [CrossRef]
- ⁴ Hann, M. M.; Leach, A. R.; Green, D. V. Em *Computational Chemistry, Molecular Complexity and Screening Set Design*, Oprea T. I., ed; Wiley-VCH: Weinheim, 2005, cap 3.
- ⁵ Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935. [CrossRef] [PubMed]
- ⁶ Klebe, G. *Drug Discov. Today: Technol.* **2004**, *1*, 225. [CrossRef]
- ⁷ Owens J. *Nat. Rev. Drug. Discov.* **2007**, *6*, 187. [CrossRef]
- ⁸ Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. C. *Nat. Biotechnol.* **2007**, *25*, 71. [CrossRef] [PubMed]
- ⁹ Marshall, G. R. Em *Introduction to chemoinformatics in drug discovery – A personal view*; Oprea, T. I.; ed.; WILEY-VCH: Weinheim, 2004. cap. 1.
- ¹⁰ Alonso, H.; Bliznyuk, A. A.; Gready, J. E. *Med. Res. Rev.* **2006**, *26*, 531. [CrossRef] [PubMed]
- ¹¹ Henzler, A. M.; Rarey, M. *Mol. Inf.* **2010**, *29*, 164. [CrossRef]
- ¹² Talevi, A.; Bruno-Blanch, L. E. *Lat. Am. J. Pharm.* **2009**, *28*, 141. [Link]
- ¹³ Kroemer, R. T. *Curr. Protein Pept. Sci.* **2007**, *8*, 312. [CrossRef] [PubMed]
- ¹⁴ Rarey, M.; Lemmen, C.; Matter, H. Em *Algorithmic Engines in Virtual Screening*; Oprea, T. I., ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2005, cap. 4.
- ¹⁵ Vyas, V.; Jain, A.; Jain, A.; Gupta, A. *Sci. Pharm.* **2008**, *76*, 333. [Link]
- ¹⁶ Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177. [CrossRef] [PubMed]
- ¹⁷ Anderson, A. C. *Chem. Biol.* **2003**, *10*, 787. [CrossRef] [PubMed]
- ¹⁸ Crivori, P.; Poggesi, I. *Eur J. Med. Chem.* **2006**, *41*, 795. [CrossRef] [PubMed]
- ¹⁹ Taft, C. A.; Da Silva, C. H. T. P.; *Current Methods in Medicinal Chemistry and Biological Physics*, 1a. ed., Research Singpost: Kerala, 2008.
- ²⁰ Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discov.* **2007**, *6*, 211. [CrossRef] [PubMed]
- ²¹ Eckert, H.; Vogt, I.; Bajorath, J. *J. Chem. Inf. Model.* **2006**, *46*, 1623. [CrossRef] [PubMed]
- ²² Muegge, I. *Mini. Rev. Med. Chem.* **2008**, *8*, 927. [CrossRef] [PubMed]
- ²³ Shoichet, B. K. *Nature* **2004**, *432*, 862. [CrossRef] [PubMed]
- ²⁴ Grabowski, K.; Baringhaus, K. H.; Schneider, G. *Nat. Prod. Rep.* **2008**, *25*, 892. [CrossRef] [PubMed]
- ²⁵ Patrick, G. L.; *An Introduction to Medicinal Chemistry*, 4a. ed., Oxford University Press: Oxford, 2009.
- ²⁶ Schneider, G. *Nat. Rev. Drug Discov.* **2010**, *9*, 273. [CrossRef] [PubMed]

- ²⁷ Drewry, D. H.; Macarron, R. *Curr. Opin. Chem. Biol.* **2010**, *14*, 289. [[CrossRef](#)] [[PubMed](#)]
- ²⁸ Brown, R. D.; Hassan, M.; Waldman, M. J. *Mol. Graph. Model.* **2000**, *18*, 427. [[CrossRef](#)] [[PubMed](#)]
- ²⁹ Lyne, P. D. *Drug Discov. Today* **2002**, *7*, 1047. [[CrossRef](#)] [[PubMed](#)]
- ³⁰ Walter, W. P.; Murcko, M. A. *Adv. Drug Deliv. Rev.* **2002**, *54*, 255. [[CrossRef](#)] [[PubMed](#)]
- ³¹ Di, L.; Kerns, E. H. *Curr. Opin. Chem. Biol.* **2003**, *7*, 402. [[CrossRef](#)] [[PubMed](#)]
- ³² Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3. [[CrossRef](#)] [[PubMed](#)]
- ³³ Lumley, J. A. *QSAR Comb. Sci.* **2005**, *24*, 1066. [[CrossRef](#)]
- ³⁴ Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G. *ChemMedChem* **2008**, *3*, 435. [[CrossRef](#)] [[PubMed](#)]
- ³⁵ Eddershaw, P. J.; Beresford, A. P.; Bayliss, M. K.; *Drug Discov. Today* **2000**, *5*, 409. [[CrossRef](#)] [[PubMed](#)]
- ³⁶ Selick, H. E.; Beresford, A. P.; Tarbit, M. H. *Drug Discov. Today* **2002**, *7*, 109. [[CrossRef](#)] [[PubMed](#)]
- ³⁷ Sheppard, G. S.; Bouska, J. J. *Drug Discov. Today: Therapeutic Strategies* **2005**, *2*, 343. [[Link](#)]
- ³⁸ Schneider, G.; Böhm, H.-J. *Drug. Discov. Today* **2002**, *7*, 64. [[CrossRef](#)] [[PubMed](#)]
- ³⁹ Cavasotto, C. N.; Orry, A. J. W. *Curr. Top. Med. Chem.* **2007**, *7*, 1006. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁰ Honorio, K. M.; Montanari, C. A.; Andricopulo, A. D. Em: *Advances and applications of structure-based approaches in drug Discovery*; Taft, C. A.; Silva, C. H. T. P., eds.; Research Signpost: Kerala, 2008, cap 5.
- ⁴¹ Zavodszky, M. I.; Rohatgi, A.; Van Voorst, J. R.; Yan, H.; Kuhn, L. A. *J. Mol. Recognit.* **2009**, *22*, 280. [[CrossRef](#)] [[PubMed](#)]
- ⁴² Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151. [[CrossRef](#)] [[PubMed](#)]
- ⁴³ Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Mulrroy, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁴ DeLano Scientific LLC. PyMOL. Version 0.99. **2006**.
- ⁴⁵ Kramer, B.; Rarey, M.; Lengauer, T. *Proteins* **1999**, *37*, 228. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁶ Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁷ Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁸ 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. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁹ Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1. [[CrossRef](#)] [[PubMed](#)]
- ⁵⁰ Waszkowycz, B. *Drug Discov. Today* **2008**, *13*, 219. [[CrossRef](#)] [[PubMed](#)]
- ⁵¹ Dias, R.; de Azevedo, W. F. *Curr. Drug. Targets* **2008**, *9*, 1040. [[CrossRef](#)] [[PubMed](#)]
- ⁵² Dugan, N.; Erkoç, Ş. *Algorithms* **2009**, *2*, 410. [[CrossRef](#)]
- ⁵³ Rarey, M.; Lemmen, C.; Matter, H. Em *Algorithmic Engines in Virtual Screening*; Oprea, T. I., ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2005, cap. 4.
- ⁵⁴ Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. *Expert Opin. Drug. Discov.* **2010**, *5*, 597. [[CrossRef](#)] [[PubMed](#)]
- ⁵⁵ Morris, G. M.; Huey, R.; Olson, A. J. *Curr. Protoc. Bioinformatic.* **2008**, *24*, 1. [[CrossRef](#)]
- ⁵⁶ Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455. [[CrossRef](#)] [[PubMed](#)]

- ⁵⁷ Lengauer, T.; *Bioinformatics: From Genomes to Drugs*, 1a ed., Wiley-VCH: Germany, 2001.
- ⁵⁸ Vieth, M.; Hirst, J. D.; Brooks, C. L. *J. Comput. Aided Mol. Des.* **1998**, *12*, 563. [[CrossRef](#)] [[PubMed](#)]
- ⁵⁹ Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biom.* **2003**, *32*, 335. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁰ CCG *The Molecular Operating Environment (MOE)*, Canada, 1996.
- ⁶¹ Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1. [[CrossRef](#)] [[PubMed](#)]
- ⁶² Clark, D. E.; Jones, G.; Willett, P.; Kenny, P. W.; Glen, R. C. *J. Chem. Inf. Comp. Sci.* **1994**, *34*, 197. [[CrossRef](#)]
- ⁶³ Moon, J. B.; Howe, W. J. *Proteins* **1991**, *11*, 314. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁴ Xu, J.; Hagler, A. *Molecules* **2002**, *7*, 566. [[CrossRef](#)]
- ⁶⁵ Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. *J. Comput. Aided Mol. Des.* **1994**, *8*, 565. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁶ Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *J. Theor. Biol.* **1983**, *104*, 359. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁷ Lagorce, D.; Pencheva, T.; Villoutreix, B. O.; Miteva, M. A. *BMC Chem. Biol.* **2009**, *9*. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁸ Maiorov, V.; Sheridan, R. P. *J. Chem. Inf. Model.* **2005**, *45*, 1017. [[CrossRef](#)] [[PubMed](#)]
- ⁶⁹ Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. *J. Comput. Aided Mol. Des.* **1994**, *8*, 153. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁰ Breda, A.; Basso, L. A.; Santos, D. S.; de Azevedo Jr, W. F. *Curr. Comput Aided Drug. Des.* **2008**, *4*, 265. [[CrossRef](#)]
- ⁷¹ OpenEye Scientific Software. FRED: Fast Rigid Exhaustive Docking. V. 2.2.5. **2009**.
- ⁷² McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. *Biopolymers* **2003**, *68*, 76. [[CrossRef](#)] [[PubMed](#)]
- ⁷³ SYBYL. SYBYL-X 1.2, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA. 2010.
- ⁷⁴ Jain, A. N. *J. Comput. Aided Mol. Des.* **1996**, *10*, 427. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁵ Jain, A. N. *J. Comput. Aided Mol. Des.* **2000**, *14*, 199. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁶ Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁷ Welch, W.; Ruppert, J.; Jain, A. N. *Chem. Biol.* **1996**, *3*, 449. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁸ Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727. [[CrossRef](#)] [[PubMed](#)]
- ⁷⁹ Antes, I. *Proteins* **2010**, *78*, 1084. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁰ Taufer, M.; Crowley, M.; Price, D.; Chien, A. A.; Brooks III, C. L.; *Parallel and Distributed Processing Symposium, 2004 Proceedings 18th International*, Santa Fe, EUA, 2004.
- ⁸¹ Xu, H.; Hsu, P-H.; Zhang, L.; Tsai, M-D.; Freitas, M. A. *J. Proteome Res.* **2010**, *9*, 3384. [[CrossRef](#)] [[PubMed](#)]
- ⁸² Hirst, J. D. *Curr. Opin. Drug Discov. Devel.* **1998**, *1*, 28. [[PubMed](#)]
- ⁸³ Schulz-Gasch, T.; Stahl, M. *J. Mol. Model.* **2003**, *9*, 47. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁴ Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Phys. Chem. Chem. Phys.* **2006**, *08*, 5166. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁵ McGovern, S. L.; Shoichet, B. K. *J. Med. Chem.* **2003**, *46*, 2895. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁶ Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213. [[PubMed](#)] [[Link](#)]
- ⁸⁷ Jenwitheesuk, E.; Horst, J. A.; Rivas, K. L.; Van Voorhis, W. C.; Samudrala, R. *Trends Pharmacol. Sci.* **2008**, *29*, 62. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁸ Broughton, H. B. *J. Mol. Graph. Model.* **2000**, *18*, 247. [[CrossRef](#)] [[PubMed](#)]
- ⁸⁹ 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. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁰ Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. *J. Mol. Biol.* **1997**, *266*, 424. [[CrossRef](#)]
- ⁹¹ Totrov, M.; Abagyan, R. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178. [[CrossRef](#)] [[PubMed](#)]
- ⁹² Geppert, H.; Vogt, M.; Bajorath, J. *J. Chem. Inf. Model.* **2010**, *50*, 205. [[CrossRef](#)] [[PubMed](#)]
- ⁹³ Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *J. Chem. Inf. Model.* **2009**, *49*, 2333. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁴ Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. *J. Med. Chem.* **2010**, *53*, 539. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁵ Tanrikulu, Y.; Proschak, E.; Werner, T.; Geppert, T.; Todoroff, N.; Klenner, A.; Kottke, T.; Sander, K.; Schneider, E.; Seifert, R.; Stark, H.; Clark, T.; Schneider, G. *ChemMedChem* **2009**, *4*, 820. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁶ Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. *Drug Discov. Today* **2008**, *13*, 23. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁷ Giganti, D.; Guillemain, H.; Spadoni, J-L.; Nilges, M.; Zagury, J-F.; Montes, M. *J. Chem. Inf. Model.* **2010**, *50*, 992. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁸ Ekins, S.; Mestres, J.; Testa, B. *Br. J. Pharmacol.* **2007**, *152*, 9. [[CrossRef](#)] [[PubMed](#)]
- ⁹⁹ Schneider, G.; Böhm, H-J. *Drug Discov. Today* **2002**, *7*, 64. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁰ Eckert, H.; Bajorath, J. *Drug Discov. Today* **2007**, *12*, 225. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰¹ Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. *Curr. Med. Chem.* **2008**, *15*, 107. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰² Langer, T.; Wolber, G. *Drug Discov. Today: Technol.* **2004**, *1*, 203. [[CrossRef](#)]
- ¹⁰³ McInnes, C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494. [[CrossRef](#)]
- ¹⁰⁴ Sun, H. *Curr. Med. Chem.* **2008**, *15*, 1018. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁵ Rella, M.; Rushworth, C. A.; Guy, J. L.; Turner, A. J.; Langer, T.; Jackson, R. M. *J. Chem. Inf. Model.* **2006**, *46*, 708. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁶ Spitzer, G. M.; Heiss, M.; Mangold, M.; Markt, P.; Kirchmair, J.; Wolber, G.; Liedl, K. R. *J. Chem. Inf. Model.* **2010**, *50*, 1241. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁷ Chen, Z.; Li, H-L.; Zhang, Q-J.; Bao, X-G.; Yu, K-Q.; Luo, X-M.; Zhu, W-L.; Jiang, H-L. *Acta Pharmacol. Sin.* **2009**, *30*, 1694. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁸ Istvan, E. S.; Deisenhofer, J. *Science* **2001**, *292*, 5519. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰⁹ Steindl, T. M.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T. *J. Comput. Aided Mol. Des.* **2006**, *20*, 703. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁰ Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. *Nucleic Acids Res.* **2010**, *38*, 609. [[CrossRef](#)] [[PubMed](#)]
- ¹¹¹ Johnson, M. A.; Maggiora, G. M.; *Concepts and applications of molecular similarity*, 1a. ed., Wiley: New York, 1990.
- ¹¹² Kubinyi, H. *J. Braz. Chem. Soc.*, **2002**, *13*, 717. [[CrossRef](#)]
- ¹¹³ Willett, P. *Drug Discov. Today* **2006**, *11*, 1046. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁴ Gasteiger, J.; Engel, T.; *Cheminformatics, a textbook*. 1a. ed., Wiley-VCH: Weinheim 2003.
- ¹¹⁵ Baldi P.; Nasr, R. *J. Chem. Inf. Model.* **2010**, *50*, 1205. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁶ Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. *J. Mol. Graph. Model.* **2010**, *29*, 157. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁷ Salim, N.; Holliday, J.; Willett, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁸ Whittle, M.; Gillet, V. J.; Willett, P.; Loesel, J. *J. Chem. Inf. Model.* **2006**, *46*, 2206. [[CrossRef](#)] [[PubMed](#)]
- ¹¹⁹ Zhang, Q.; Muegge, I. *J. Med. Chem.* **2006**, *49*, 1536. [[CrossRef](#)] [[PubMed](#)]

- ¹²⁰ Abdo, A.; Salim, N. *QSAR Comb. Sci.* **2009**, *28*, 654. [[Link](#)]
- ¹²¹ Hert, J.; Willett, P.; Wilton, D. J. *J. Chem. Inf. Model.* **2006**, *46*, 462. [[CrossRef](#)] [[PubMed](#)]
- ¹²² Jennings, A.; Tennant, M. *J. Chem. Inf. Model.* **2007**, *47*, 1829. [[CrossRef](#)] [[PubMed](#)]
- ¹²³ Tresadern, G.; Bemporad, D.; Howe, T. J. *Mol. Graph. Model.* **2009**, *27*, 860. [[CrossRef](#)] [[PubMed](#)]
- ¹²⁴ Grant, J. A.; Gallardo, M. A.; Pickup, B. T. *J. Comput. Chem.* **1996**, *17*, 1653. [[CrossRef](#)]
- ¹²⁵ Nicholls, A.; MacCuish, N. E.; MacCuish, J. D. *J. Comput. Aided Mol. Des.* **2004**, *18*, 451. [[CrossRef](#)] [[PubMed](#)]
- ¹²⁶ Hu, L.; Chen, G.H.; Chau, R. M-W. *J. Mol. Graph. Model.* **2006**, *24*, 244. [[CrossRef](#)] [[PubMed](#)]
- ¹²⁷ Han, L. Y.; Ma, X. H.; Lin, H. H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z. R.; Cao, Z. W.; Ji, Z. L.; Chen, Y. Z. *J. Mol. Graph. Model.* **2008**, *26*, 1276. [[CrossRef](#)] [[PubMed](#)]
- ¹²⁸ Hsieh, J. H.; Wang, X. S.; Teotico, D.; Golbraikh, A.; Tropsha, A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 593. [[CrossRef](#)] [[PubMed](#)]
- ¹²⁹ Todeschini, R. *Molecular descriptors and chemometrics: A powerful combined tool for pharmaceutical, toxicological and environmental problems*. Disponível em <http://www.iamc-online.org/tutorials/T2_moleculardescriptors_chemom.pdf>. Acesso em: 22 março de 2012.
- ¹³⁰ Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959. [[CrossRef](#)] [[PubMed](#)]
- ¹³¹ Khanfar, M. A.; El Sayed, K. A. *Eur. J. Med. Chem.* **2010**, *45*, 5397. [[CrossRef](#)] [[PubMed](#)]
- ¹³² Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *24*, 4130. [[CrossRef](#)] [[PubMed](#)]
- ¹³³ Thondorf, I.; Voigt, V.; Schäfer, S.; Gebauer, S.; Zebisch, K.; Laug, L.; Brandsch, M. *Bioorg. Med. Chem.* **2011**, *19*, 6409. [[CrossRef](#)] [[PubMed](#)]
- ¹³⁴ Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, *119*, 10509. [[CrossRef](#)]
- ¹³⁵ Hopfinger, A. J.; Reaka, A.; Venkatarangan. P.; Duca, J. S.; Wang, S. *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 1151. [[CrossRef](#)]
- ¹³⁶ Myint, K. Z.; Xie, X-Q. *Int. J. Mol. Sci.* **2010**, *11*, 3846. [[CrossRef](#)] [[PubMed](#)]
- ¹³⁷ Vedani, A.; Dobler, M.; Lill, M. A. *Basic Clin. Pharmacol. Toxicol.* **2006**, *99*, 195. [[CrossRef](#)] [[PubMed](#)]
- ¹³⁸ Lill, M. A.; Dobler, M.; Vedani, A. *Curr. Comput. Aided Drug. Des.* **2005**, *1*, 307. [[CrossRef](#)]
- ¹³⁹ Polanski, J. *Curr. Med. Chem.* **2009**, *16*, 3243. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴⁰ Melville, J. L.; Burke, E. K.; Hirst, J. D. *Comb. Chem High T. Scr.* **2009**, *12*, 332. [[Link](#)]
- ¹⁴¹ Duch, W.; Swaminathan, K.; Meller, J. *Curr. Pharm. Des.* **2007**, *13*, 1497. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴² Fogel, G. B. *Brief Bioinform.* **2008**, *9*, 307. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴³ Li, H.; Yap, C. W.; Ung, C. Y.; Xue, Y.; Li, Z. R.; Han, L. Y.; Lin, H. H.; Chen, Y. Z. *J. Pharm. Sci.* **2007**, *96*, 2838. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴⁴ Burton, J.; Ijjaali, I.; Petitet, F.; Michel, A.; Vercauteren, D. P. *Comb. Chem. High T. Scr.* **2009**, *12*, 369. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴⁵ Liu, P.; Long, W. *Int. J. Mol. Sci.* **2009**, *10*, 1978. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴⁶ Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴⁷ Krovat, E. M.; Steindl, T.; Langer, T. *Curr. Comput. Aided Drug. Des.* **2005**, *1*, 93. [[CrossRef](#)]
- ¹⁴⁸ Khedkar, S. A.; Malde, A. K.; Coutinho, E. C. *Indian J. Pharm. Sci.* **2008**, *68*, 689. [[Link](#)]

- ¹⁴⁹ Muegge, I.; Oloff, S. *Drug Discov. Today: Technol.* **2006**, *3*, 405. [[CrossRef](#)]
- ¹⁵⁰ Rao, H.; Yang, G.; Tan, N.; Li, P.; Li, Z.; Li, X. *QSAR Comb. Sci.* **2009**, *28*, 1346. [[CrossRef](#)]
- ¹⁵¹ Habibi-Yangjeh, A. *Monatsh. Chem.* **2009**, *140*, 523. [[CrossRef](#)]
- ¹⁵² Karelson, M.; Dobchev, D.; Tamm, T.; Tulp, I.; Janes, J.; Tamm, K.; Lomaka, A.; Savechenko, D.; Karelson, G. *ARKIVOC* **2008**, *16*, 38. [[Link](#)]
- ¹⁵³ Worachartcheewan, A.; Nantasenamat, C.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. *Eur. J. Med. Chem.* **2009**, *44*, 1664. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵⁴ Chamjangali, M. A. *Chem. Biol. Drug. Des.* **2009**, *73*, 456. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵⁵ Fernandes, M. B.; Scotti, M. T.; Ferreira, M. J. P.; Emerenciano, V. P. *Eur. J. Med. Chem.* **2008**, *43*, 2197. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵⁶ Hristozov, D.; Oprea, T. I.; Gasteiger, J. J. *Chem. Inf. Model.* **2007**, *47*, 2044. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵⁷ Kaiser, D.; Terfloth, L.; Kopp, S.; Schulz, J.; de Laet, R.; Chiba, P.; Ecker, G. F.; Gasteiger, J. J. *Med. Chem.* **2007**, *50*, 1698. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵⁸ Nascimento, R. F. F.; Alcântara, E. H.; Kampel, M.; Stech, J. L.; Novo, E. M. L. M.; Fonseca, L. M. G. *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, Brasil, 2009.
- ¹⁵⁹ Vapnik, V. N.; *The nature of statistical learning theory*, 1a. ed., Springer-Verlag: New York, 1995.
- ¹⁶⁰ Cristianini, N.; Shawe-Taylor, J.; *An Introduction To Support Vector Machines (and other kernel-based learning methods)*, 1a. ed., Cambridge University Press, 2000.
- ¹⁶¹ Gonçalves, A. R., Máquina de Vetores Suporte. Acesso em 21 de março de 2012. [[Link](#)]
- ¹⁶² Smola, A. J.; Bartlett, P.; Schuurmans, C. *Introduction to large margin classifiers*; Scholkopf, B.; Schuurmans, D.; Smola, A. J.; Schölkopf, B., eds.; MIT Press: Cambridge, 2000, cap. 1.
- ¹⁶³ Andreola, R.; Haertel, V.; *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, Brasil, 2009.
- ¹⁶⁴ Lorena, A. C.; de Carvalho, A. C. P. L. F., Introdução às Maquinas de Vetores Suporte. Disponível em: <http://www.icmc.usp.br/~biblio/BIBLIOTEC_A/rel_tec/RT_192.pdf>. Acesso em: 22 março 2012.
- ¹⁶⁵ Lima, A. R. G.; *Dissertação de Mestrado*, Universidade Federal do Ceará, Brasil, 2002. [[Link](#)]
- ¹⁶⁶ Zien, A.; Rätsch, G.; Mika, S.; Schölkopf, B.; Lengauer, T.; Müller, K-R. *Bioinformatics* **2000**, *16*, 799. [[CrossRef](#)] [[PubMed](#)]
- ¹⁶⁷ Sung, A. H.; Mukkamala, S.; *Proceedings of the 2003 International Symposium on Applications and the Internet Technology*, Orlando, EUA, 2003.
- ¹⁶⁸ Ding, C. H.; Dubchak, I. *Bioinformatics* **2001**, *17*, 349. [[CrossRef](#)] [[PubMed](#)]
- ¹⁶⁹ Cong, Y.; Yang, X-G.; Lv, W.; Xue Y. *J. Mol. Graph. Model.* **2009**, *28*, 236. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷⁰ Vasanathan, P.; Taboureau, O.; Oostenbrink, C.; Vermeulen, N. P. E.; Olsen, L.; Jørgensen, F. S. *Drug Metab. Dispos.* **2009**, *37*, 658. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷¹ Liew, C. Y.; Ma, X. H.; Liu, X.; Yap, C. W. *J. Chem. Inf. Model.* **2009**, *49*, 877. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷² Mahé, P.; Ralaivola, L.; Stoven, V.; Vert, J. P. *J. Chem. Inf. Model.* **2006**, *46*, 2003. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷³ Yang, X. G.; Lv, W.; Chen, Y. Z.; Xue, Y. *J. Comput. Chem* **2010**, *31*, 1249. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷⁴ Mballo, C.; Makarenkov, V. *Comb. Chem. High. T. Scr.* **2010**, *13*, 430. [[PubMed](#)]
- ¹⁷⁵ Agarwal, S.; Dugar, D.; Sengupta, S. *J. Chem. Inf. Model.* **2010**, *50*, 716. [[CrossRef](#)] [[PubMed](#)]

- ¹⁷⁶ Nasonov, A. F. *Russ. J. Gen. Chem.* **2010**, *80*, 2541. [[CrossRef](#)]
- ¹⁷⁷ Liu, X. H.; Song, H. Y.; Zhang, J. X.; Han, B. C.; Wei, X. N.; Ma, X. H.; Cui, W. K.; Chen, Y. Z. *Molecular Informatics* **2010**, *29*, 407. [[CrossRef](#)]
- ¹⁷⁸ Hao, M.; Li, Y.; Wang, Y.; Zhang, S. *Anal. Chim. Acta.* **2011**, *690*, 53. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷⁹ Kinnings, S. L.; Liu, N.; Tonge, P. J.; Jackson, R. M.; Xie, L.; Bourne, P. E. *J. Chem. Inf. Model.* **2011**, *51*, 408. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁰ Rajwade, R.; Pande, R.; Mishra, K. P.; Kumar, A.; Pandey, B. N. *QSAR Comb. Sci.* **2009**, *28*, 1500. [[CrossRef](#)]
- ¹⁸¹ Du-Cuny, L.; Song, Z.; Moses, S.; Powis, G.; Mash, E. A.; Meuillet, E. J.; Zhang S. *Bioorg. Med. Chem.* **2009**, *17*, 6983. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸² Peterson, Y. K.; Wang, X. S.; Casey, P. J.; Tropsha, A. *J. Med. Chem.* **2009**, *52*, 4210. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸³ Jensen, B. F.; Vind, C.; Padkjar, S. B.; Brockhoff, P. B.; Refsgaard, H. H. F. *J. Med. Chem.* **2007**, *50*, 501. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁴ Cedeño, W.; Agrafiotis, D. K. *J. Comput. Aided Mol. Des.* **2003**, *17*, 255. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁵ Miller, D. W. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 68. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁶ Shen, M.; Béguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. *J. Med. Chem.* **2004**, *47*, 2356. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁷ Oloff, S.; Zhang, S.; Sukumar, N.; Breneman, C.; Tropsha, A. *J. Chem. Inf. Model.* **2006**, *46*, 844. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁸ Zhang, S. X.; Golbraikh, A.; Tropsha, A. *J. Med. Chem.* **2006**, *49*, 2713. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸⁹ Wang, X. S.; Tang, H.; Golbraikh, A.; Tropsha, A. *J. Chem. Inf. Model.* **2008**, *48*, 997. [[CrossRef](#)] [[PubMed](#)]
- ¹⁹⁰ Gunturi, S.; Archana, K.; Khandelwal, A.; Narayanan, R. *QSAR Comb. Sci.* **2008**, *27*, 1305. [[CrossRef](#)]