UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

ANACLETO SILVA DE SOUZA

Relações quantitativas entre a estrutura e atividade para uma série de
antichagásicos derivados do fenarimol

São Carlos
2015

ANACLETO SILVA DE SOUZA

Relações quantitativas entre a estrutura e atividade para uma série de antichagásicos derivados do fenarimol

Dissertação apresentada ao Programa de Pós- Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de Concentração: Física Aplicada

Opção: Física Biomolecular

Orientador: Prof. Dr. Adriano Defini Andricopulo

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2015

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Silva de Souza, Anacleto

Relações quantitativas entre a estrutura e atividade para uma série de antichagásicos derivados do fenarimol / Anacleto Silva de Souza; orientador Adriano Defini Andricopulo - versão corrigida -- São Carlos, 2015.

104 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2015.

1. Química medicinal . 2. Doença de Chagas. 3. QSAR. 4. HQSAR. I. Defini Andricopulo, Adriano, orient. II. Título.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me concedido a oportunidade de concluir meu trabalho de mestrado e pela minha vida.

À minha família e a Janaína Angélica Borges, por me apoiarem, sempre.

Ao Prof. Dr. Adriano D. Andricopulo, por ter me concedido a oportunidade de desenvolver esse trabalho.

Ao Dr. Leonardo Luiz Gomes Ferreira pelo auxílio no desenvolvimento desta dissertação.

A todos os meus amigos e colegas de trabalho, em especial, Leonardo Luiz Gomes Ferreira, Ricardo Nascimento dos Santos e César Maschio Fioravanti, por terem me auxiliado e me ajudado durante o projeto.

Aos professores, funcionários e colegas do Laboratório de Química Medicinal e Computacional do Instituto de Física de São Carlos pela colaboração e suporte.

Ao Instituto de Física de São Carlos, pela oportunidade de realizar o curso de mestrado.

A FAPESP, CNPq e CAPES pelo auxílio financeiro para a realização deste trabalho. Em especial, a CAPES, pela bolsa concedida que possibilitou a minha dedicação aos estudos.

RESUMO

SILVA DE SOUZA, A. **Relações quantitativas entre a estrutura e atividade para uma série de antichagásicos derivados do fenarimol**. 2015. 104 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

A doença de Chagas é uma das doenças tropicais negligenciadas prioritárias para a Organização Mundial da Saúde (OMS). Endêmica na América Latina, é considerada atualmente um problema de saúde mundial, afetando diversos países na América do Norte, Europa, Ásia e Oceania. Estima-se que 8-10 milhões de pessoas estejam infectadas com o protozoário *Trypanosoma cruzi*. O tratamento disponível é limitado a dois fármacos que apresentam baixa eficácia e sérios efeitos adversos, o que torna evidente a urgência por novas alternativas terapêuticas. Esta dissertação de mestrado tem como objetivo o desenvolvimento de estudos das relações quantitativas entre a estrutura e atividade (QSAR, na sigla inglesa para *quantitative structure-activity relationships*) para duas séries de derivados do fenarimol com atividade anti-*T. cruzi*. Os compostos selecionados possuem considerável potência *in vitro* contra o parasita, além de atividade *in vivo* em modelos experimentais da doença, o que caracteriza o seu potencial para estudos em química medicinal. Neste contexto, estratégias de planejamento de fármacos foram utilizadas para a geração de modelos de QSAR preditivos. Foram utilizados os métodos holograma QSAR (HQSAR, na sigla inglesa para *hologram QSAR*); análise comparativa dos campos moleculares (CoMFA, na sigla inglesa para *comparative molecular field analysis*); e análise comparativa dos índices de similaridade estrutural (CoMSIA, na sigla inglesa para *comparative molecular similarity indices*). Os modelos gerados possuem alta capacidade de correlação interna e de predição externa, indicando também um conjunto de características estruturais responsáveis pela atividade anti-*T. cruzi*. Os resultados apresentados neste trabalho são úteis no planejamento de novos derivados do fenarimol com propriedades antichagásicas.

Palavras-chave: Química medicinal. Doença de Chagas. QSAR. HQSAR.

ABSTRACT

SILVA DE SOUZA, A. **Quantitative structure-activity relationships for a series of antichagasic fenarimol derivatives.** 2015. 104 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

Chagas' disease is considered by the World Health Organization (WHO) as one of the top neglected tropical diseases. Endemic in Latin America, it is currently a global public health problem, affecting several countries in North America, Europe, Asia and Oceania. The disease, caused by the protozoan parasite *Trypanosoma cruzi*, affects around 8-10 million people worldwide. The available treatment is limited to two drugs that present low efficacy and severe side effects, highlighting the urgent need for new therapeutic options. This master's dissertation focuses on the development of quantitative structure-activity relationships (QSAR) models for two series of fenarimol derivatives with activity against *T. cruzi*. The selected compounds exhibit substantial *in vitro* potency against the parasite, as well as *in vivo* activity in experimental models of the disease, which point out their potential for further studies in medicinal chemistry. In this context, drug design strategies were applied for the generation of predictive QSAR models. The methods employed were hologram QSAR (HQSAR); comparative molecular field analysis (CoMFA); and comparative molecular similarity indices (CoMSIA). The models possess high internal and external consistency, also indicating a set of structural features related to their anti-*T. cruzi* activity. The results reported herein are useful for the design of new fenarimol derivatives as new antichagasic agents.

Keywords: Medicinal chemistry. Chagas' disease. QSAR. HQSAR.

LISTA DE FIGURAS

Figura 1 –	Geração de um modelo HQSAR, a partir da fragmentação de uma estrutura molecular bidimensional.	26
Figura 2 –	Esquema geral do processo de construção de um modelo CoMFA.	28
Figura 3 –	Distribuição global da doença de Chagas.	30
Figura 4 –	Medicamentos utilizados no tratamento da doença de Chagas.	31
Figura 5 –	Antifúngicos contendo o anel triazol com atividade <i>in vitro</i> e <i>in vivo</i> contra o <i>Trypanosoma cruzi</i> .	32
Figura 6 –	Fungicidas ativos contra o <i>T. cruzi</i> .	33
Figura 7 –	A) Estrutura bidimensional do inibidor EPL-BS967. B) Estrutura cristalográfica da enzima CYP51 de <i>Trypanosoma cruzi</i> em complexo com o inibidor EPL-BS967 (PDB 3ZG3, resolução de 2,9 Å). A CYP51 está representada em <i>cartoon</i> . Os átomos de carbono do inibidor estão coloridos em ciano, da protoporfirina em magenta e o átomo de ferro em verde.	34
Figura 8 –	Mapa de similaridade estrutural para o conjunto 1 e distribuição das moléculas por faixa de atividade.	59
Figura 9 –	Alinhamento molecular para o conjunto de dados 1.	60
Figura 10 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo HQSAR (conjunto de dados 1).	63
Figura 11 –	Mapas de contribuição de HQSAR para alguns compostos do conjunto de dados 1.	64
Figura 12 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo CoMFA (conjunto de dados 1).	67
Figura 13 –	Mapas de contorno CoMFA para algumas moléculas do conjunto de dados 1.	68

Figura 14 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo CoMSIA (conjunto de dados 1).	71
Figura 15 –	Mapas de contorno estereoquímicos e eletrostáticos gerados pelo modelo CoMSIA para algumas moléculas do conjunto de dados 1.	72
Figura 16 –	Mapas de contorno para doadores e aceptores de interações de hidrogênio e para grupos hidrofóbicos, gerados pelo modelo CoMSIA para algumas moléculas do conjunto de dados 1.	73
Figura 17 –	Mapa de similaridade estrutural para o conjunto 2 e distribuição das moléculas por faixa de atividade.	80
Figura 18 –	Alinhamento molecular para o conjunto de dados 2.	81
Figura 19 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo de HQSAR (conjunto de dados 2).	84
Figura 20 –	Mapas de contribuição de HQSAR para alguns compostos do conjunto de dados 2.	85
Figura 21 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo CoMFA (conjunto de dados 2).	87
Figura 22 –	Mapas de contorno CoMFA para algumas moléculas do conjunto de dados 2.	88
Figura 23 –	Valores experimentais <i>versus</i> preditos de pIC ₅₀ para o melhor modelo CoMSIA (conjunto de dados 2).	90
Figura 24 –	Mapas de contorno estereoquímicos e eletrostáticos para o melhor modelo CoMSIA (conjunto de dados 2).	91
Figura 25 –	Mapas de contorno para doadores e receptores de interações de hidrogênio, e para grupos hidrofóbicos, gerados pelo modelo CoMSIA, para algumas moléculas do conjunto de dados 2.	92
Figura 26 –	Estratégias a serem exploradas em trabalhos de otimização molecular para esta série de compostos. A) Composto 41 do conjunto de dados 1. B) Composto 18 do conjunto de dados 2.	96

LISTA DE TABELAS

Tabela 1 –	Parâmetros utilizados para distinguir os fragmentos moleculares.	25
Tabela 2 –	Estruturas e valores de pIC_{50} para o conjunto de dados 1.	49
Tabela 3 –	Resultados de HQSAR para o conjunto de dados 1, utilizando diferentes parâmetros de distinção de fragmentos, tamanho de fragmento de 4 a 7 átomos e método de validação LOO.	61
Tabela 4 –	Resultados gerados para diferentes tamanhos de fragmento (modelo A/B/C/H/Ch).	61
Tabela 5 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo HQSAR - conjunto de dados 1).	62
Tabela 6 –	Modelos CoMFA obtidos através do método SDC de focalização de região (conjunto de dados 1).	65
Tabela 7 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA - conjunto de dados 1).	66
Tabela 8 –	Resultados CoMSIA obtidos através da variação do valor do fator de atenuação (conjunto de dados 1).	69
Tabela 9 –	Resultados CoMSIA obtidos através da variação do valor de SDC (conjunto de dados 1).	69
Tabela 10 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA - conjunto de dados 1).	70
Tabela 11 –	Estruturas e valores de pIC_{50} para o conjunto de dados 2.	74
Tabela 12 –	Resultados de HQSAR para o conjunto de dados 2, utilizando diferentes parâmetros de distinção de fragmentos, tamanho de fragmento de 4 a 7 átomos e método de validação LOO.	82
Tabela 13 –	Resultados gerados para diferentes tamanhos de fragmento (modelo A/B/C/H/Ch).	82

Tabela 14 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo de HQSAR - conjunto de dados 2).	83
Tabela 15 –	Modelos CoMFA obtidos através do método SDC de focalização de região (conjunto de dados 2).	86
Tabela 16 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA - conjunto de dados 2).	86
Tabela 17 –	Resultados CoMSIA obtidos através da variação do valor do fator de atenuação (conjunto de dados 2).	89
Tabela 18 –	Resultados CoMSIA obtidos através da variação do valor de SDC (conjunto de dados 2).	89
Tabela 19 –	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA - conjunto de dados 2).	90

LISTA DE ABREVIATURAS E SIGLAS

OMS	Organização Mundial da Saúde
P&D	Pesquisa e desenvolvimento
NCE	<i>New Chemical Entity</i>
SBDD	<i>Structure-based drug design</i>
LBDD	<i>Ligand-based drug design</i>
VS	<i>Virtual screening</i>
QSAR	<i>Quantitative structure-activity relationships</i>
$\log P$	Logarítmo do coeficiente de partição
HQSAR	<i>Hologram quantitative structure-activity relationships</i>
A	Átomos
B	Ligações
C	Conectividade
H	Hidrogênios
Ch	Quiralidade
DA	Doador e receptor
PLS	<i>Partial least squares</i>
2D	Bidimensional
3D	Tridimensional
CoMFA	<i>Comparative molecular field analysis</i>
CoMSIA	<i>Comparative molecular similarity indices</i>
DTN	Doença tropical negligenciada
DNDi	<i>Drugs for Neglected Diseases Initiative</i>

<i>T. cruzi</i>	<i>Trypanosoma cruzi</i>
IC ₅₀	Concentração de um composto necessária para inibir em 50% uma determinada resposta biológica
pIC ₅₀	-log IC ₅₀
PCA	<i>Principal component analysis</i>
SDC	<i>Standard deviation coefficient</i>
LOO	<i>Leave-one-out</i>
LMO	<i>Leave-many-out</i>
q^2	Coeficiente de correlação com validação cruzada
r^2	Coeficiente de correlação sem validação cruzada
SEE	<i>Standard error of estimate</i>
HL	<i>Hologram length</i>
N	Número de componentes
SEP	<i>Standard error of prediction</i>
r^2_{pred}	Coeficiente de predição

SUMÁRIO

1 Introdução	19
1.1 Planejamento de Fármacos	21
1.2 Relações Quantitativas entre a Estrutura e Atividade	22
1.2.1 <i>Holograma QSAR</i>	24
1.2.2 <i>Análise Comparativa dos Campos Moleculares</i>	26
1.2.3 <i>Análise Comparativa dos Índices de Similaridade Molecular</i>	28
1.3 Doença de Chagas	29
1.4 Alternativas para o Tratamento da doença de Chagas	31
2 Objetivos	35
3 Materiais e Métodos	39
3.1 Recursos Computacionais	41
3.2 Conjuntos de Dados	41
3.3 Estruturas 3D e Alinhamento Molecular	42
3.4 Análise de Componentes Principais e Seleção das Séries Treinamento e Teste	43
3.5 Modelagem de HQSAR	43
3.6 Modelagem CoMFA	44
3.7 Modelagem CoMSIA	44
3.8 Validação dos Modelos de QSAR	45
4 Resultados e Discussão	47
4.1 Conjunto de Dados 1	49
4.1.1 <i>Diversidade Estrutural e Distribuição da Propriedade Alvo</i>	49
4.1.2 <i>Alinhamento Molecular</i>	60

4.1.3 Modelos de HQSAR – Conjunto 1	60
4.1.4 Modelos CoMFA – Conjunto 1	65
4.1.5 Modelos CoMSIA – Conjunto 1	68
4.2 Conjunto de Dados 2	73
4.2.1 Diversidade Estrutural e Distribuição da Propriedade Alvo	73
4.2.2 Alinhamento Molecular	80
4.2.3 Modelos de HQSAR – Conjunto 2	81
4.2.4 Modelos CoMFA – Conjunto 2	86
4.2.5 Modelos CoMSIA – Conjunto 2	89
5 Conclusões e Perspectivas	93
Referências	97

Introdução

1 Introdução

1.1 Planejamento de Fármacos

O processo de pesquisa e desenvolvimento (P&D) de novos fármacos faz uso, cada vez mais, de estratégias modernas de química medicinal. Essa área se dedica a identificação de compostos que possuem atividade biológica, ao estudo de seus mecanismos de ação em nível molecular, e ao planejamento de novas entidades químicas (NCEs, na sigla inglesa para *new chemical entities*) candidatas a fármacos.¹

Os avanços alcançados nas ciências químicas, biológicas, médicas e afins, contribuíram decisivamente para a evolução dos métodos utilizados em química medicinal.² A compreensão de processos moleculares essenciais para a manutenção da vida resultou em um novo paradigma, no qual o estudo e a predição da modulação de sistemas biológicos adquiriram grande importância³. Neste cenário, ferramentas computacionais têm sido úteis em uma diversidade de etapas, incluindo a identificação de alvos moleculares; a predição de propriedades farmacocinéticas e farmacodinâmicas; bem como a simulação de interações entre fármacos e seus receptores biológicos. Estas abordagens *in silico* têm contribuído efetivamente no processo de planejamento de novos fármacos, apresentando constante evolução nas últimas décadas.³

As estratégias em química medicinal podem ser classificadas em duas grandes classes: (i) os métodos de planejamento de fármacos baseados na estrutura do receptor (SBDD, na sigla inglesa para *structure-based drug design*); e, (ii) os métodos de planejamento baseados na estrutura do ligante (LBDD, na sigla inglesa para *ligand-based drug design*).⁴⁻⁵ Os métodos de SBDD utilizam informações estruturais de macromoléculas que constituem os alvos biológicos para ação dos fármacos. A compreensão de fenômenos de reconhecimento entre uma molécula de baixo peso molecular (ligante) e o seu respectivo alvo molecular (receptor) é o principal objetivo das técnicas SBDD.⁶ A química medicinal utiliza várias abordagens de SBDD, como a triagem virtual (VS, na sigla inglesa para *virtual*

screening), a docagem e a dinâmica molecular, entre outras, que levaram a descoberta de inúmeras moléculas com alta afinidade e potência.

A estratégia de LBDD utiliza informações de moléculas com atividade biológica determinada em modelos experimentais bem definidos.⁷ Características moleculares são utilizadas para a construção de modelos computacionais, usualmente representados como modelos quantitativos de relações entre a estrutura e atividade (QSAR, na sigla inglesa para *quantitative structure-activity relationships*).⁸ Estes modelos são aplicados para orientar o planejamento de novas moléculas e predizer sua atividade no sistema experimental investigado.⁹ Tanto os métodos SBDD, quanto os métodos LBDD, são amplamente utilizados, e têm contribuído de forma sólida para o desenvolvimento de candidatos a novos agentes terapêuticos.¹⁰⁻¹¹

A indústria farmacêutica se destaca mundialmente pelos elevados investimentos em inovação.¹² O desenvolvimento de um novo fármaco é um processo longo e complexo, que leva mais de uma década em P&D, atingindo custos da ordem de 1 bilhão de dólares.¹³ O número de novos fármacos lançados no mercado nos últimos anos não tem acompanhado o crescimento nos investimentos em P&D.¹⁴ O grande desafio para as maiores companhias farmacêuticas está em converter a inovação científica e tecnológica em programas de P&D mais eficientes, e em última análise, em fármacos inovadores.

1.2 Relações Quantitativas entre a Estrutura e Atividade

Dentre as estratégias mais importantes de LBDD, encontram-se os métodos de QSAR,¹⁵ que se baseiam no princípio de que a variação em uma determinada propriedade biológica resulta de diferenças moleculares específicas de ligantes. Seu objetivo é o desenvolvimento de modelos matemáticos capazes de correlacionar as características moleculares e as propriedades biológicas. De forma geral, as propriedades empregadas nestes estudos são parâmetros farmacodinâmicos e farmacocinéticos, determinados experimentalmente.¹⁶ Os modelos originados podem ser aplicados para predizer a atividade de novos compostos, bem como para

identificar as propriedades moleculares mais relevantes em questão. Assim, é possível planejar as variações estruturais/moleculares a serem exploradas no planejamento de novos compostos.¹⁷

A atividade biológica, denominada variável dependente ou propriedade alvo, consiste, na maioria dos estudos de QSAR, em medidas de potência ou afinidade (farmacodinâmica), ou em propriedades farmacocinéticas, como absorção e biodisponibilidade.¹⁸ Essas propriedades são obtidas por meio de modelos experimentais específicos (e.g., ensaios *in vivo* ou *in vitro*).¹⁹⁻²⁰ Esses dados devem ser obtidos de forma padronizada e validada para que possam ser úteis em estudos de QSAR.

As propriedades moleculares, denominadas descritores ou variáveis independentes, são obtidas por meio de cálculos computacionais ou de medidas experimentais. Uma ampla variedade de descritores moleculares é utilizada nos estudos de QSAR. Descritores como o coeficiente de partição ($\log P$), área de superfície polar, peso molecular, número de ligações rotacionáveis e de aceptores e doadores de ligações de hidrogênio, podem estar diretamente correlacionados às propriedades alvo (e.g., farmacodinâmicas e farmacocinéticas).²¹ Nos métodos de impressão digital molecular, os descritores são notações numéricas que codificam a estrutura molecular bidimensional.²² Existem também os métodos que consideram a conformação tridimensional das moléculas para gerar valores de energia de interação intermolecular, que por sua vez, são utilizados como descritores.²³

O desenvolvimento de modelos de QSAR consistentes e com boa capacidade preditiva, depende de diversos fatores que devem ser considerados: (i) seleção do método estatístico mais adequado; (ii) o conjunto de dados deve conter moléculas com boa variação estrutural e que atuem pelo mesmo mecanismo, com dados biológicos padronizados e validados, apresentando boa distribuição; (iii) os conjuntos treinamento e teste devem ser representativos de todo o intervalo da propriedade alvo; (iv) é essencial a adoção de métodos robustos de validação interna e externa.²⁴⁻²⁵

1.2.1 Holograma QSAR

O método de holograma QSAR (HQSAR, na sigla inglesa para *hologram quantitative structure-activity relationships*) converte a estrutura bidimensional das moléculas do conjunto de dados em uma sequência numérica denominada impressão digital molecular. No método de HQSAR, a impressão digital é chamada de holograma molecular, que difere da impressão digital tradicional, por não ser uma notação do tipo binária.²⁶ Como a propriedade alvo é função da estrutura e o holograma contém informações sobre a estrutura, a propriedade alvo se torna função do holograma.²⁷

O algoritmo do método de HQSAR fragmenta a estrutura bidimensional das moléculas do conjunto de dados. Neste processo são gerados fragmentos moleculares lineares, ramificados e cíclicos, e que podem sobrepor-se uns aos outros. Os fragmentos são distribuídos ao longo do holograma, de forma que cada posição da sequência contenha a identidade e o número de fragmentos destinados a esta posição.

Diversos parâmetros afetam significativamente a qualidade dos modelos HQSAR. Variáveis como o comprimento do holograma, o tamanho dos fragmentos moleculares e os critérios utilizados para distinguir os fragmentos, devem ser exploradas durante a modelagem. O comprimento do holograma se refere ao número de posições disponíveis para que os fragmentos sejam distribuídos. O número mínimo e máximo de átomos que cada fragmento pode conter é definido pelo tamanho dos fragmentos. As características utilizadas para distinguir os inúmeros fragmentos são determinadas pelos parâmetros de distinção de fragmentos.

A tabela 1 resume as características que determinam a distinção entre os fragmentos moleculares. Geralmente, várias combinações dessas características são utilizadas e avaliadas quanto à robustez estatística dos modelos gerados.

Tabela 1 – Parâmetros utilizados para distinguir os fragmentos moleculares.

Parâmetros de distinção de fragmentos	Definição
Átomos (A)	Os tipos de átomos são utilizados na distinção entre os fragmentos
Ligações (B)	Diferencia os fragmentos de acordo com as ligações: simples, duplas, triplas e aromáticas
Conectividade (C)	Refere-se ao estado de hibridização dos átomos contidos no fragmento
Hidrogênios (H)	Distingue os fragmentos segundo a presença e número de átomos de hidrogênio
Quiralidade (Ch)	Utiliza a presença de centros quirais para distinguir os fragmentos moleculares
Doador e Aceptor (DA)	Diferencia os fragmentos de acordo com a presença de átomos aceptores e doadores de ligação de hidrogênio

Fonte: Elaborada pelo autor

Uma vez gerados os hologramas moleculares, o método estatístico de regressão linear por mínimos quadrados parciais (PLS, na sigla inglesa para *partial least squares*), gera um modelo que correlaciona a propriedade alvo com as informações codificadas nos hologramas (figura 1).²⁸ Após a realização dos procedimentos de validação, o modelo é empregado na predição da propriedade alvo de novas moléculas.

Os modelos HQSAR também são utilizados na identificação dos fragmentos moleculares que contribuem para a variação da propriedade alvo. Com esta finalidade são gerados mapas de contribuição que utilizam um código de cores para indicar a importância de cada fragmento para a atividade biológica.²⁹

Figura 1 – Geração de um modelo HQSAR, a partir da fragmentação de uma estrutura molecular bidimensional.

Fonte: Elaborada pelo autor

1.2.2 Análise Comparativa dos Campos Moleculares

A análise comparativa dos campos moleculares (CoMFA, na sigla inglesa para *comparative molecular field analysis*) é o método de QSAR 3D mais amplamente utilizado. O método CoMFA se baseia no princípio de que as interações intermoleculares determinam a propriedade alvo.³⁰ Como esse processo de reconhecimento molecular tem natureza tridimensional, o método exige a construção de conformações 3D das moléculas a serem estudadas. Os descritores moleculares são os valores da energia de interação entre os átomos das moléculas investigadas e uma sonda virtual. Esta sonda consiste em grupo funcional ou um átomo que apresenta propriedades estereoquímicas e eletrostáticas definidas.³¹ A variação dos valores de energia é o fator que explica a variação da propriedade alvo. Os tipos de interação utilizados no método CoMFA são as interações eletrostáticas e as

interações de van der Waals, descritas pelos potenciais de Coulomb e Lennard-Jones, respectivamente.³²⁻³³

O método CoMFA requer o alinhamento das moléculas do conjunto de dados no espaço cartesiano. Todas as estruturas devem apresentar conformações semelhantes e a mesma orientação, para que a energia de interação calculada possa ser comparada. Dentre os diferentes métodos de alinhamento molecular, destacam-se os que se baseiam em uma subestrutura comum a todo o conjunto, os que utilizam a docagem molecular, ou ainda os que utilizam grupos farmacofóricos.³⁴⁻³⁶

Uma grade tridimensional é gerada em torno das moléculas do conjunto de dados e a energia de interação entre a sonda e a superfície molecular é calculada em cada ponto da grade. A cada ponto são atribuídos dois valores de energia, um para a interação eletrostática e outro para a interação de van der Waals. Em seguida, o método PLS gera um modelo estatístico que correlaciona a propriedade biológica com os valores de energia. A figura 2 ilustra a geração de um modelo CoMFA.

Figura 2 – Esquema geral do processo de construção de um modelo CoMFA.

Fonte: Elaborada pelo autor

Através da construção de mapas de contorno 3D, o método CoMFA permite a análise visual de características moleculares que determinam variações significativas na propriedade alvo. Esses mapas revelam regiões em torno da superfície molecular em que alterações das características eletrostáticas e estereoquímicas podem resultar em moléculas com propriedades otimizadas.

1.2.3 Análise Comparativa dos Índices de Similaridade Molecular

A análise comparativa dos índices de similaridade molecular (CoMSIA, na sigla inglesa para *comparative molecular similarity indices*) é uma extensão do método CoMFA.³⁷ O método CoMSIA considera os campos moleculares eletrostáticos, estereoquímicos, hidrofóbicos, bem como a presença de grupos doadores e aceptores de interações de hidrogênio. Estas características individualizam importantes contribuições para a interação fármaco-receptor. No método CoMSIA, os descritores moleculares são os índices de similaridade, calculados entre a sonda virtual e as moléculas do conjunto de dados, em cada ponto de interseção da grade. Uma função gaussiana é utilizada na determinação

dos índices de similaridade. Neste método, um conjunto de mapas de contorno indica regiões do espaço ocupado pelas moléculas, onde a presença de grupos que apresentam determinada característica físico-química é favorável ou desfavorável para a propriedade alvo.

1.3 Doença de Chagas

A doença de Chagas é uma das doenças tropicais negligenciadas (DTNs), definidas pela Organização Mundial da Saúde (OMS). As DTNs são um grupo de 17 doenças que possui como característica comum maior prevalência em regiões pobres e remotas do planeta, de clima tropical e subtropical.³⁸ Embora essas doenças afetem mais de um bilhão de pessoas, os investimentos em P&D são extremamente modestos quando comparados às doenças prioritárias para a indústria farmacêutica, que afetam principalmente os países desenvolvidos. A transmissão das DTNs é favorecida pela falta de condições mínimas aceitáveis de saneamento e habitação, e agravada por alterações ambientais e proliferação de insetos que atuam como vetores destas doenças.³⁹

A doença de Chagas é uma doença infecciosa crônica causada pelo protozoário *Trypanosoma cruzi*. A doença afeta aproximadamente 8-10 milhões de pessoas, sendo endêmica na América Latina, principalmente, em países como Brasil, Argentina e México.⁴⁰ Devido à migração populacional, a doença de Chagas é considerada atualmente um problema de saúde pública mundial, sendo encontrada também em países desenvolvidos (e.g., América do Norte, Europa, Ásia e Oceania).⁴¹ A figura 3 mostra a distribuição mundial da doença e o número estimado de casos em cada região do planeta.

Figura 3 – Distribuição global da doença de Chagas.

Fonte: Adaptado de: WHO, 2013.⁴⁰

A doença de Chagas é transmitida por insetos hematófagos e afeta mais de 150 espécies de animais, além do homem. Os triatomíneos *Triatoma infestans*, *Rhodnius prolixus*, e *Triatoma dimidiata* são as principais espécies de vetores que transmitem a doença para o ser humano.⁴² A transmissão ocorre quando os insetos eliminam o parasita junto às fezes ao se alimentar de sangue. O parasita também pode ser transmitido através de transfusão sanguínea, transplante de órgãos e via transplacentária. Raramente ocorrem casos de transmissão através de alimentos contaminados com o parasita e através de acidentes laboratoriais.⁴²

Apesar de ter sido descrita há mais de 100 anos, o tratamento da doença de Chagas é extremamente limitado e ineficaz.⁴³ Apenas dois fármacos, descobertos na década de 1970, são utilizados no tratamento da doença: o nifurtimox, comercializado sob o nome de Lampit[®] pela Bayer, e o benzonidazol, vendido como Rochagan[®] pela Roche. A figura 4 mostra a estrutura destes dois medicamentos, que apresentam baixa eficácia (10 a 20% de cura na fase crônica da doença e 60 a 80% na fase aguda) e efeitos adversos severos (náusea, perda de peso, dor abdominal, anorexia e alterações hematológicas).⁴⁴ De acordo com a Iniciativa

Fármacos para Doenças Negligenciadas (DNDi, na sigla inglesa para *Drugs for Neglected Diseases Initiative*), é urgente o desenvolvimento de novos fármacos que possuam um perfil adequado, caracterizado por: (i) eficácia nas duas fases clínicas da doença (aguda e crônica); (ii) ação contra todas as variantes do parasita em todas as regiões geográficas; (iii) potência contra cepas resistentes; (iv) toxicidade reduzida; (v) alta biodisponibilidade oral; e, (v) duração máxima de 30 dias para o tratamento, preferencialmente, por via oral.⁴⁵

Figura 4 – Medicamentos utilizados no tratamento da doença de Chagas.

Fonte: Elaborada pelo autor.

1.4 Alternativas para o Tratamento da Doença de Chagas

O uso de fármacos dirigidos ao tratamento de outras doenças, é uma das estratégias mais exploradas em DTNs.⁴⁶ Moléculas que possuem perfil farmacocinético e toxicidade definidos em humanos, são muito atrativas para uma área que não é prioritária em programas de P&D do setor farmacêutico. A eflornitina, um fármaco inicialmente desenvolvido para o tratamento do câncer, compõe o arsenal terapêutico para a tripanossomíase Africana humana.⁴⁷ Os antifúngicos cetoconazol e fluconazol são utilizados no tratamento da leishmaniose.⁴⁸ O reposicionamento de fármacos oferece inúmeras vantagens e diversos trabalhos vêm sendo desenvolvidos para a doença de Chagas.⁴⁹

Derivados azólicos são utilizados como antifúngicos desde a década de 1970.⁵⁰ Esses compostos vêm sendo investigados como agentes tripanocidas desde a década de 1980, quando foi constatada atividade inibitória dos antifúngicos

miconazol e econazol sobre o crescimento do *T. cruzi*.⁵¹ O antifúngico posaconazol e o pró-fármaco do ravuconazol, denominado E1224 (figura 5), demonstraram atividade em modelos *in vitro* e *in vivo* da doença de Chagas, e chegaram a entrar em fase clínica de desenvolvimento.⁵²

Figura 5 – Antifúngicos contendo o anel triazol com atividade *in vitro* e *in vivo* contra o *Trypanosoma cruzi*.

Fonte: Elaborada pelo autor.

Estes resultados impulsionaram investigações sobre a atividade anti-*T. cruzi* de agroquímicos utilizados como antifúngicos em lavouras.⁵³ Os agroquímicos e os fármacos anti-infecciosos possuem em comum o objetivo de controlar o crescimento de microrganismos, sem causar danos à saúde humana. A aprovação de agentes agroquímicos requer avaliações de toxicidade, à medida que muitos destes compostos são utilizados em alimentos que são consumidos por humanos. O reposicionamento de agentes agroquímicos para uso medicinal é, portanto, uma estratégia útil a ser explorada na área de DTNs.⁵⁴

Em estudos conduzidos com agroquímicos, foram identificadas dezenas de compostos com atividade *in vitro* sobre o *T. cruzi*. Dentre as moléculas mais potentes, a maioria consiste em derivados azólicos.⁵⁴ O fungicida fenarimol, originalmente desenvolvido como agroquímico na década de 1970, destaca-se como um derivado não azólico, em que o grupo imidazol ou triazol é substituído por um

anel pirimidínico.⁵⁵ Estas moléculas são extremamente potentes, possuindo atividade *in vitro* na faixa de concentração nanomolar ($IC_{50} = 350$ nM, para o fenarimol). Além disso, o fenarimol apresentou atividade contra diversas espécies de *Leishmania*, que causam a leishmaniose. Estes estudos demonstram a utilidade do reposicionamento de fármacos e de agroquímicos no desenvolvimento de novos agentes tripanocidas. A figura 6 mostra a estrutura destes fungicidas, identificados como potentes agentes anti-*T. cruzi*.

Figura 6 – Fungicidas ativos contra o *T. cruzi*.

Fonte: Elaborada pelo autor.

Estudos recentes de cristalografia de raios-X indicam que os derivados do fenarimol agem como inibidores da enzima 14 α -esterol demetilase do *Trypanosoma cruzi* (CYP51), uma enzima envolvida na biossíntese de esteróis.⁵⁶ A figura 7 mostra a estrutura da CYP51 de *T. cruzi* em complexo com o derivado piridínico EPL-BS967 (PDB 3ZG3, resolução de 2,9 Å).⁵⁷

Figura 7 – A) Estrutura bidimensional do inibidor EPL-BS967. B) Estrutura cristalográfica da enzima CYP51 de *Trypanosoma cruzi* em complexo com o inibidor EPL-BS967 (PDB 3ZG3, resolução de 2,9 Å). A CYP51 está representada em *cartoon*. Os átomos de carbono do inibidor estão coloridos em ciano, da protoporfirina em magenta e o átomo de ferro em verde.

Fonte: Elaborado pelo autor.

Objetivos

2 Objetivos

O objetivo geral desta dissertação de mestrado é a geração de modelos de QSAR capazes de prever a atividade anti-*T. cruzi in vitro* de compostos estruturalmente relacionados ao fenarimol. Como objetivos específicos, destacam-se:

- ✓ Organização de conjuntos de dados de compostos com atividade anti-*T. cruzi* para a geração de modelos de QSAR.
- ✓ Desenvolvimento de modelos de QSAR 2D e QSAR 3D, empregando as técnicas HQSAR, CoMFA e CoMSIA.
- ✓ Avaliação da consistência estatística dos modelos de QSAR, incluindo sua capacidade de correlação interna para o conjunto treinamento, e de predição externa para o conjunto teste.
- ✓ Avaliação das principais características estruturais por meio dos mapas 2D de contribuição (HQSAR) e 3D de contorno (CoMFA e CoMSIA).

Materiais e Métodos

3 Materiais e Métodos

3.1 Recursos Computacionais

A geração e a análise dos modelos de QSAR foram realizadas no Laboratório de Química Medicinal e Computacional do Instituto de Física de São Carlos – Universidade de São Paulo, conforme descrito em trabalhos anteriores.⁵⁸⁻⁶⁰ Foram utilizadas estações computacionais Dell XPS que operam através do sistema Linux CentOS. A plataforma de modelagem molecular SYBYL-X 2.0 (Tripos, St. Louis, EUA) foi utilizada na geração e análise dos modelos de QSAR.

3.2 Conjuntos de Dados

Neste trabalho foram utilizados dois conjuntos de dados para a construção dos modelos. Estes conjuntos, denominados 1 e 2, foram selecionados da literatura,^{55,61} organizados e usados para a geração de modelos de QSAR 2D e 3D. Para esta finalidade, foram empregados o método 2D HQSAR e os métodos 3D CoMFA e CoMSIA.

O conjunto 1 consiste em 77 moléculas inspiradas na estrutura do fenarimol.⁵⁵ A atividade destes compostos foi medida em condições experimentais padronizadas e descrita em valores de IC_{50} , que se referem a concentração necessária de composto para inibir em 50% a reprodução *in vitro* do *T. cruzi*. Os valores de IC_{50} variam entre 1 nM e 27 μ M, um intervalo de potência de 27.000 vezes. Estes valores foram convertidos em $-\log IC_{50}$ (pIC_{50}) e assim empregados na modelagem de QSAR.

O conjunto 2 consiste em 52 derivados do fenarimol, com atividade tripanocida *in vitro*.⁶¹ Os ensaios biológicos foram conduzidos nas mesmas condições experimentais empregadas para o conjunto 1. Para o conjunto 2, os valores de IC_{50} variam entre 5 nM e 1 μ M, um fator de potência de 200 vezes. Estes

valores também foram convertidos em pIC_{50} para o trabalho de modelagem de QSAR.

Os dois conjuntos de dados foram selecionados porque reúnem as condições necessárias para a geração de modelos de QSAR. Principalmente, diversidade estrutural adequada; boa distribuição dos valores da propriedade alvo; e padronização dos ensaios biológicos. Ademais, esta classe de compostos possui relevante atividade anti-*T. cruzi in vitro* e baixa toxicidade. Estudos recentes indicam que estes inibidores da enzima 14 α -esterol demetilase (CYP51) possuem eficácia comparável à do posaconazol e superior à do benzonidazol, em modelos *in vivo* da doença de Chagas.⁶² Essa classe possui, portanto, alto potencial para o desenvolvimento de novos candidatos a agentes antichagásicos.

3.3 Estruturas 3D e Alinhamento Molecular

As estruturas 3D das moléculas dos conjuntos de dados foram originadas utilizando-se os parâmetros geométricos e atômicos padrões do programa SYBYL-X 2.0. A minimização energética foi conduzida com o método Powell, usando-se os seguintes parâmetros: método do gradiente conjugado; critério de convergência de 0,005 kcal/molÅ; campo de força Tripos; cargas Gasteiger-Huckel; e constante dielétrica de 80.⁶³⁻⁶⁵

A obtenção de modelos de QSAR 3D requer o alinhamento estrutural do conjunto de dados. Este alinhamento 3D foi realizado com o método Distill Rigid,⁶⁶ o qual identifica uma subestrutura comum entre as moléculas a serem alinhadas e utiliza os valores dos ângulos diedros desta subestrutura como referência para o alinhamento das demais moléculas. A conformação de mínima energia da molécula mais potente de cada conjunto (compostos **61** e **31**, para os conjuntos 1 e 2, respectivamente) foi utilizada como modelo para as demais moléculas.

3.4 Análise de Componentes Principais e Seleção das Séries Treinamento e Teste

A seleção dos conjuntos treinamento e teste é um procedimento crítico para a obtenção de modelos de QSAR preditivos. Estes conjuntos devem incluir moléculas representativas de todo o espaço químico (diversidade estrutural) e biológico (propriedade alvo) ocupado pelas moléculas investigadas. Com este objetivo, o método de análise de componentes principais (PCA, na sigla inglesa para *principal component analysis*) foi empregado para avaliar a similaridade estrutural entre os compostos e a sua distribuição em relação à propriedade alvo.⁶⁷ Impressões digitais moleculares geradas pelo método UNITY foram construídas a partir da estrutura bidimensional das moléculas, e empregadas como descritores moleculares.⁶⁸ O resultado da PCA é um gráfico de similaridade estrutural em que cada ponto representa uma molécula do conjunto. Mapas de similaridade estrutural foram gerados para os dois conjuntos de dados investigados com a finalidade de selecionar conjuntos treinamento e teste representativos da diversidade estudada.

3.5 Modelagem de HQSAR

Modelos de HQSAR foram originados para os conjuntos de dados 1 e 2. Para a geração dos descritores moleculares foram utilizados os comprimentos de holograma padrão: 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, e 401. Com o objetivo de distinguir os diferentes fragmentos gerados pelo algoritmo de HQSAR, diversas combinações de parâmetros de distinção de fragmentos foram utilizadas. O efeito da variação do tamanho dos fragmentos sobre os modelos também foi investigado.

3.6 Modelagem CoMFA

Modelos CoMFA foram gerados para os conjuntos de dados 1 e 2. Cada conjunto foi inserido em uma caixa tridimensional reticulada com espaçamento de 2 Å entre os vértices de cada retículo e 4 Å de extensão além da superfície das moléculas. A energia de interação entre os átomos e a sonda virtual (átomo de carbono sp^3 com carga líquida +1) foi calculada com o campo de força Tripos, e os potenciais de Lennard-Jones e Coulomb, para as interações de van der Waals e eletrostáticas, respectivamente. Como valor de corte para os dois potenciais foi utilizado o valor padrão de 30 kcal/mol. O recurso de focalização de região foi empregado para refinar e melhorar a resolução dos modelos. Para este fim o método de ajuste dos coeficientes de desvio padrão (SDC, na sigla inglesa para *standard deviation coefficient*) foi aplicado.⁶⁹

3.7 Modelagem CoMSIA

Modelos CoMSIA foram gerados para os conjuntos 1 e 2. As moléculas de cada conjunto foram inseridas em uma grade idêntica à utilizada na modelagem CoMFA. Uma função gaussiana, que considera a distância entre os átomos de cada molécula e a sonda virtual (átomo de carbono sp^3 com carga líquida +1) foi utilizada para calcular os índices de similaridade em cada ponto da grade. Diferentes valores do fator de atenuação foram empregados na implementação da função. Todas as características físico-químicas disponíveis no módulo CoMSIA foram utilizadas no cálculo dos índices de similaridade: características hidrofóbicas, estereoquímicas e eletrostáticas, e interações de hidrogênio. O método SDC de focalização de região também foi aplicado aos modelos CoMSIA.

3.8 Validação dos Modelos de QSAR

Todos os modelos de QSAR gerados foram submetidos aos procedimentos de validação interna (avaliação da capacidade de correlação para o conjunto treinamento) e externa (avaliação da capacidade de predição para o conjunto teste). Apenas os conjuntos treinamento foram utilizados para gerar os modelos de QSAR. A validação interna foi realizada com o método “deixe um de fora” (LOO, na sigla inglesa para *leave-one-out*).⁷⁰ Este procedimento determina o número ótimo de componentes do modelo e o coeficiente de correlação com validação cruzada (q^2). Em seguida foi realizada uma análise PLS completa, empregando o número ótimo de componentes para gerar o coeficiente de correlação sem validação cruzada (r^2). Para avaliar a estabilidade dos modelos, foi também aplicado o método “deixe vários de fora” (LMO, na sigla inglesa para *leave-many-out*).⁷⁰ Uma vez estabelecida a consistência interna dos modelos, sua habilidade em prever a atividade de moléculas de um conjunto externo (teste) foi investigada. Assim, os melhores modelos foram empregados na predição dos valores de pIC_{50} para dois conjuntos teste.

Resultados e Discussão

4 Resultados e Discussão

4.1 Conjunto de Dados 1

4.1.1 Diversidade Estrutural e Distribuição da Propriedade Alvo

Modelos de HQSAR, CoMFA e CoMSIA foram desenvolvidos para um conjunto de 77 derivados do fenarimol que possui atividade anti-*T. cruzi in vitro*. As estruturas das moléculas que compõem o conjunto de dados e os respectivos valores de pIC_{50} são listados na tabela 2.

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
1		6,46	2		6,66
3		7,62	4		6,85
5		6,08	6		6,84

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
7		6,71	8		6,57
9		7,02	10		7,77
11		7,46	12		7,00
13		7,51	14		8,10

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
15		6,36	16		4,64
17		4,57	18		5,15
19		7,05	20		7,30
21		7,05	22		6,22
23		5,97	24		6,17

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
25		6,64	26		7,10
27		6,43	28		7,77
29		7,22	30		7,40
31		6,35	32		7,09

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
33		8,70	34		8,40
35		7,15	36		7,27
37		7,51	38		7,70
39		6,92	40		7,15

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
41		8,10	42		6,96
43		7,15	44		7,92
45		8,70	46		8,00
47		7,52	48		7,60

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
49		7,15	50		6,92
51		7,52	52		8,22
53		8,70	54		7,80
55		8,7	56		7,68

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
57		7,59	58		7,59
59		7,59	60		8,52
61		9,00	62		7,80
63		7,82	64		7,80

(Continua)

(Continuação)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
65		7,89	66		8,30
67		7,11	68		7,70
69		6,70	70		6,85
71		6,59	72		7,68

(Continua)

(Conclusão)

Tabela 2 – Estruturas e valores de pIC_{50} para o conjunto de dados 1.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
73		6,31	74		5,82
75		6,52	76		7,22
77		8,10			

Figuras geradas no programa *ChemBioDraw 14.0*.

Fonte: Elaborada pelo autor.

Este conjunto de dados foi submetido a uma análise PCA com a finalidade de se avaliar a similaridade estrutural entre os compostos e a sua distribuição em relação aos valores de pIC_{50} . Como resultado foi gerado um mapa de similaridade estrutural em que cada ponto representa uma molécula do conjunto. A distância entre os pontos é inversamente proporcional à similaridade entre os compostos e o código de cores identifica as diferentes faixas de atividade do conjunto. A partir do

mapa de similaridade estrutural, foram selecionadas a série treinamento, utilizada para a construção dos modelos, e a série teste, empregada no processo de validação externa. A figura 8 mostra o mapa de similaridade estrutural para o conjunto 1 e a distribuição das moléculas de acordo com os valores de pIC_{50} . O gráfico evidencia uma grande concentração de moléculas na faixa de pIC_{50} entre 7 e 8. O intervalo compreendido entre 4 e 6 contém o menor número de moléculas. A seleção dos conjuntos treinamento e teste foi realizada de modo a incluir em ambos os conjuntos moléculas estruturalmente diversas e que representem a distribuição da propriedade alvo. Desta maneira foram selecionadas as seguintes moléculas para o conjunto teste: **2, 6, 14, 17, 21, 23, 25, 33, 34, 46, 52, 54, 62, 64, 70, 72 e 76**.

Figura 8 – Mapa de similaridade estrutural para o conjunto 1 e distribuição das moléculas por faixa de atividade.

Fonte: Elaborada pelo autor.

4.1.2 Alinhamento Molecular

O conjunto de dados 1 (tabela 2) foi submetido a um processo de alinhamento molecular, necessário para a obtenção dos modelos CoMFA e CoMSIA. As estruturas 3D das moléculas, construídas conforme descrito na seção 3.3, foram utilizadas. O alinhamento, obtido através do método Distill Rigid, foi realizado com base na subestrutura comum a todos os compostos, utilizando como referência o composto mais potente (**61**). Todas as moléculas alinhadas do conjunto 1 são ilustradas na figura 9.

Figura 9 – Alinhamento molecular para o conjunto de dados 1.

Fonte: Elaborada pelo autor.

4.1.3 Modelos de HQSAR – Conjunto 1

Modelos de HQSAR foram desenvolvidos para o conjunto de dados 1. Todos os comprimentos de holograma disponíveis (53 a 401) foram utilizados na construção dos modelos. Inicialmente, foram gerados modelos com as seguintes combinações dos parâmetros de distinção de fragmentos: A/B/C, A/C/DA,

A/B/C/H/Ch e A/B/C/H/DA. Foram empregados o tamanho de fragmento padrão de 4 a 7 átomos e o método de validação *LOO*. A tabela 3 mostra os resultados obtidos para as diferentes combinações utilizadas.

Tabela 3 – Resultados de HQSAR para o conjunto de dados 1, utilizando diferentes parâmetros de distinção de fragmentos, tamanho de fragmento de 4 a 7 átomos e método de validação *LOO*.

Distinção de Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
A/B/C	0,58	0,76	0,41	97	3
A/C/DA	0,54	0,80	0,38	257	4
A/B/C/H/Ch	0,60	0,81	0,37	97	5
A/B/C/H/DA	0,59	0,80	0,39	97	5

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *HL*, Comprimento do holograma; *N*, Número de componentes.

Fonte: Elaborada pelo autor.

Como pode ser observado, o melhor modelo foi obtido com a combinação A/B/C/H/Ch ($q^2 = 0,60$; $r^2 = 0,81$; e *SEE* = 0,37). O tamanho dos fragmentos gerados também interfere nos resultados. O efeito da variação deste parâmetro foi avaliado para o melhor modelo da tabela 3. A tabela 4 lista os resultados obtidos para os tamanhos de fragmento: 3–7, 3–9, 4–7, 4–8, 4–9, e 4–10.

Tabela 4 – Resultados gerados para diferentes tamanhos de fragmento (modelo A/B/C/H/Ch).

Tamanho de Fragmento	Parâmetros Estatísticos					
	q^2	r^2	SEE	HL	N	r^2_{pred}
3-7	0,59	0,86	0,32	401	6	0,64
3-9	0,64	0,89	0,29	401	6	0,62
4-7	0,60	0,81	0,37	97	5	0,61
4-8	0,60	0,84	0,35	53	5	0,52
4-9	0,63	0,89	0,29	401	6	0,62
4-10	0,57	0,90	0,28	257	7	0,66

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *HL*, Comprimento do holograma; *N*, Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

Os resultados obtidos através da variação do tamanho de fragmentos são comparáveis àqueles obtidos com o tamanho de fragmento padrão. Um discreto aumento nos valores de r^2 e uma diminuição nos valores do erro são observados. Os

modelos também foram avaliados quanto a sua estabilidade frente à variação de dados. Com este objetivo, o conjunto de dados foi dividido aleatoriamente em cinco grupos e a sua capacidade de predição foi avaliada pelo método LMO ($q^2 = 0,50$ e $0,49$ para cinco e dez grupos, respectivamente).

O processo de validação mais robusto é o que avalia a habilidade dos modelos em prever a propriedade alvo de compostos que não foram utilizados na sua construção. Uma série teste de compostos, selecionada de acordo com o mapa de similaridade estrutural (figura 8), foi utilizada no processo de validação externa do melhor modelo de HQSAR para o conjunto 1 (A/B/C/H/Ch, tabela 3). O processo de validação foi aplicado a este modelo para todos os tamanhos de fragmentos listados na tabela 4. Como pode ser observado, o modelo com maior capacidade preditiva é o que apresenta fragmentos que variam de 4 a 10 átomos ($r^2_{\text{pred}} = 0,66$). A comparação entre os valores preditos e experimentais de pIC_{50} é exibida na tabela 5 e mostrada graficamente na figura 10.

Tabela 5 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo HQSAR - conjunto de dados 1).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	6,66	6,52	0,14
6	6,84	6,64	0,20
14	8,10	7,64	0,46
17	4,57	6,00	-1,43
21	7,05	6,87	0,18
23	5,97	6,15	-0,18
25	6,64	6,48	0,16
33	8,70	7,52	1,18
34	8,40	7,94	0,47
46	8,00	7,89	0,11
52	8,22	8,13	0,09
54	7,80	7,79	0,01
62	7,80	7,27	0,53
64	7,80	7,95	-0,15
70	6,85	6,68	0,17
72	7,68	6,49	1,19
76	7,22	7,24	-0,02

^a Diferença entre os valores experimentais e preditos de pIC_{50} .

Fonte: Elaborada pelo autor.

Figura 10 – Valores experimentais *versus* preditos de pIC_{50} para o melhor modelo de HQSAR (conjunto de dados 1).

Fonte: Elaborada pelo autor.

A boa correlação entre os valores experimentais e preditos de pIC_{50} para o conjunto teste, e o valor do coeficiente de predição ($r^2_{\text{pred}} = 0,66$) para o melhor modelo de HQSAR, indicam sua utilidade para prever a atividade de novos compostos, estruturalmente relacionados ao fenarimol.

Os mapas de contribuição de HQSAR são empregados na avaliação de fragmentos que contribuem positivamente ou negativamente para a propriedade biológica. Ao passo que as contribuições negativas são representadas em vermelho, vermelho-alaranjado e laranja, as contribuições positivas são representadas em amarelo, verde-azulado e verde. A cor branca indica contribuições intermediárias ou neutras. A figura 11 ilustra os mapas de contribuição para alguns compostos representativos do conjunto de dados 1.

Figura 11 – Mapas de contribuição de HQSAR para alguns compostos do conjunto de dados 1.

Fonte: Elaborada pelo autor.

Os mapas de contribuição indicam que moléculas potentes, como os compostos **28**, **41**, **45** e **53** (pIC_{50} = 7,77; 8,10; 8,70 e 8,70, respectivamente) possuem um grande número de átomos que contribuem positivamente para a atividade biológica. Esta contribuição positiva é em grande parte atribuída à presença de um anel benzênico nestas moléculas. Em contrapartida, a molécula **22** possui átomos nas posições 3 e 4 do anel piridínico que contribuem negativamente. O mapa é compatível com a baixa potência apresentada por esta molécula (pIC_{50} = 6,22). O composto **16**, cujo carbono 4 do anel pirimidínico é assinalado como uma contribuição negativa, também apresenta baixa atividade (pIC_{50} = 4,64).

4.1.4 Modelos CoMFA – Conjunto 1

Modelos CoMFA foram desenvolvidos com o objetivo de investigar os aspectos tridimensionais que determinam a atividade dos compostos do conjunto de dados 1. A consistência interna dos modelos foi avaliada primeiramente através do método de validação cruzada LOO. Diversos modelos CoMFA foram gerados através da utilização do recurso de focalização de região. O ajuste dos coeficientes de desvio padrão (SDC) foi aplicado para a focalização. Valores de SDC de 0,3; 0,6 e 0,9 foram usados para eliminar descritores redundantes e aumentar o peso dos que contribuem de maneira mais significativa para a atividade. Os modelos gerados através da focalização de região são apresentados na tabela 6.

Tabela 6 – Modelos CoMFA obtidos através do método SDC de focalização de região (conjunto de dados 1).

SDC	q^2	SEP	r^2	SEE	N	r^2_{pred}
0,3	0,50	0,61	0,75	0,43	5	0,61
0,6	0,53	0,59	0,72	0,47	5	0,59
0,9	0,54	0,59	0,71	0,46	5	0,54

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEE , Erro padrão de estimativa; SEP , Erro padrão de predição; N , Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

Como pode ser observado, o modelo com maior capacidade preditiva para o conjunto teste ($r^2_{\text{pred}} = 0,61$) foi obtido ajustando-se os valores de SDC por um fator de 0,3. Os modelos também foram avaliados quanto a sua estabilidade frente a variação de dados. Para isso, o conjunto de dados foi dividido aleatoriamente e a sua capacidade de predição foi avaliada pelo método LMO ($q^2 = 0,40$ e $0,50$ para cinco e dez grupos, respectivamente). A diferença entre os valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (SDC = 0,3; tabela 6) é exibida na tabela 7 e ilustrada graficamente na figura 12.

Tabela 7 – Valores experimentais e preditos de plC_{50} para o conjunto teste (modelo CoMFA - conjunto de dados 1).

Composto	plC_{50} Experimental	plC_{50} Predito	Resíduo ^a
2	6,66	6,90	-0,24
6	6,84	6,56	0,27
14	8,10	7,40	0,70
17	4,57	5,34	-0,77
21	7,05	6,65	0,40
23	5,97	5,86	0,11
25	6,64	6,27	0,37
33	8,70	7,14	1,56
34	8,40	7,59	0,81
46	8,00	7,81	0,19
52	8,22	8,67	-0,45
54	7,80	7,48	0,32
62	7,80	7,76	0,04
64	7,80	8,81	-1,01
70	6,85	7,31	-0,46
72	7,68	7,39	0,29
76	7,22	6,62	0,60

^a Diferença entre os valores experimentais e preditos de plC_{50}

Fonte: Elaborada pelo autor.

Figura 12 – Valores experimentais *versus* preditos de pIC_{50} para o melhor modelo CoMFA (conjunto de dados 1).

Fonte: Elaborada pelo autor.

A correlação observada entre os valores experimentais e preditos de pIC_{50} para o conjunto teste, obtida pelo melhor modelo CoMFA, demonstra sua utilidade em prever a atividade de novos compostos relacionados ao fenarimol.

As relações entre a estrutura e atividade para o conjunto 1 foram interpretadas por meio de mapas de contorno eletrostáticos e estereoquímicos gerados pelo melhor modelo CoMFA. Os mapas de contorno indicam regiões em que alterações nas características eletrostáticas e estereoquímicas determinam variações no valor da propriedade alvo. A figura 13 mostra os mapas eletrostáticos e estereoquímicos para os compostos **16**, **22**, **24** e **41**. Os poliedros verdes e azuis indicam regiões em que a presença de grupos químicos volumosos e eletropositivos, respectivamente, é favorável para a atividade biológica.

Figura 13 – Mapas de contorno CoMFA para algumas moléculas do conjunto de dados 1.

Fonte: Elaborada pelo autor.

4.1.5 Modelos CoMSIA – Conjunto 1

Modelos CoMSIA foram desenvolvidos para o conjunto de dados 1. A consistência interna dos modelos foi avaliada através do método de validação cruzada LOO. Primeiramente, foi investigado o efeito da variação do fator de atenuação sobre os resultados (tabela 8). O melhor modelo foi obtido aplicando-se um valor de 0,7 para o fator de atenuação.

Tabela 8 – Resultados CoMSIA obtidos através da variação do valor do fator de atenuação (conjunto de dados 1).

Fator de Atenuação	q^2	SEP	r^2	SEE	N
0,3	0,16	0,81	0,93	0,24	8
0,4	0,21	0,79	0,94	0,22	8
0,5	0,24	0,77	0,95	0,21	8
0,6	0,26	0,77	0,95	0,20	8
0,7	0,27	0,75	0,95	0,20	7
0,8	0,26	0,76	0,94	0,21	7

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *SEP*, Erro padrão de predição; *N*, Número de componentes.

Fonte: Elaborada pelo autor.

Os valores obtidos para o coeficiente de correlação com validação cruzada q^2 , contudo, não foram satisfatórios. Desta forma, o método de focalização de região SDC foi utilizado para eliminar os descritores redundantes e aumentar o peso dos que contribuem de maneira mais consistente para o modelo. Os modelos gerados através da focalização de região são apresentados na tabela 9.

Tabela 9 – Resultados CoMSIA obtidos através da variação do valor de SDC (conjunto de dados 1).

SDC	q^2	SEP	r^2	SEE	N	r^2_{pred}
0,6	0,62	0,54	0,91	0,26	8	0,67
0,9	0,61	0,56	0,89	0,30	8	0,67

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *SEP*, Erro padrão de predição; *N*, Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

Como pode ser observado, os dois modelos da tabela 9 apresentam o mesmo valor de r^2_{pred} (igual a 0,67). No entanto, o modelo com valor de SDC = 0,6 é ligeiramente melhor em relação aos demais parâmetros estatísticos. A estabilidade do modelo foi avaliada através do método LMO ($q^2 = 0,54$ e 0,62, para cinco e dez grupos, respectivamente). A diferença entre os valores experimentais e preditos de pIC_{50} para o melhor modelo CoMSIA (SDC = 0,6; tabela 9) é exibida na tabela 10 e ilustrada graficamente na figura 14.

Tabela 10 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA - conjunto de dados 1).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	6,66	6,89	-0,23
6	6,84	6,57	0,26
14	8,10	7,39	0,70
17	4,57	5,33	-0,76
21	7,05	6,65	0,39
23	5,97	5,85	0,11
25	6,64	6,27	0,3
33	8,70	7,14	1,55
34	8,40	7,59	0,80
46	8,00	7,81	0,19
52	8,22	8,67	-0,45
54	7,80	7,47	0,32
62	7,80	7,76	0,03
64	7,80	8,81	-1,01
70	6,85	7,30	-0,45
72	7,68	7,39	0,29
76	7,22	6,61	0,60

^a Diferença entre os valores experimentais e preditos de pIC_{50}

Fonte: Elaborada pelo autor.

Figura 14 – Valores experimentais *versus* preditos de pIC_{50} para o melhor modelo CoMSIA (conjunto de dados 1).

Fonte: Elaborada pelo autor.

O método CoMSIA inclui nos seus mapas de contorno, características eletrostáticas, estereoquímicas, hidrofóbicas, e a presença de grupos doadores e aceptores de ligação de hidrogênio. Os mapas de contorno indicam regiões em que alterações nestas características determinam variações no valor da propriedade alvo. A figura 15 mostra os mapas CoMSIA para os compostos **16**, **22**, **24** e **41**. Os poliedros azuis indicam regiões em que a presença de grupos eletropositivos são favoráveis para a atividade, e os poliedros amarelos indicam regiões em que grupos volumosos são desfavoráveis para a atividade biológica. De acordo com o mapa, a presença de grupos volumosos como substituintes na posição *para* dos anéis (benzeno, pirimidina e piridina), é prejudicial para a atividade biológica.

Figura 15 – Mapas de contorno estereoquímicos e eletrostáticos gerados pelo modelo CoMSIA para algumas moléculas do conjunto de dados 1.

Fonte: Elaborada pelo autor.

A figura 16 mostra os mapas CoMSIA para os compostos **16**, **22**, **24** e **41**. Os poliedros coloridos em ciano indicam que a presença de grupos doadores de interações de hidrogênio é favorável para a atividade. O poliedro magenta indica que grupos aceptores de interações de hidrogênio são favoráveis para a atividade biológica. Os poliedros amarelos indicam que a presença de grupos hidrofóbicos é favorável para a atividade destas moléculas.

Figura 16 – Mapas de contorno para doadores e aceptores de interações de hidrogênio e para grupos hidrofóbicos, gerados pelo modelo CoMSIA para algumas moléculas do conjunto de dados 1.

Fonte: Elaborada pelo autor.

4.2 Conjunto de Dados 2

4.2.1 Diversidade Estrutural e Distribuição da Propriedade Alvo

Modelos de HQSAR, CoMFA e CoMSIA foram desenvolvidos para o segundo conjunto de derivados do fenarimol, constituído por 52 compostos com atividade

anti-*T. cruzi* *in vitro*. As estruturas das moléculas deste conjunto de dados e os respectivos valores de pIC_{50} são listados na tabela 11.

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
1		7,72	2		7,62
3		7,59	4		7,55
5		7,52	6		7,46
7		7,44	8		7,17

(Continua)

(Continuação)

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
9		7,15	10		7,09
11		6,77	12		6,62
13		6,60	14		6,55
15		6,29	16		8,10
17		8,05	18		8,00
19		7,92	20		7,74

(Continua)

(Continuação)

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
21		7,62	22		7,60
23		7,55	24		7,55
25		7,46	26		7,11
27		6,96	28		6,44
29		6,11	30		6,06

(Continua)

(Continuação)

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
31		8,30	32		8,22
33		8,15	34		8,10
35		8,05	36		8,00
37		7,77	38		7,70
39		7,66	40		7,62

(Continua)

(Continuação)

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
41		7,27	42		6,92
43		7,77	44		7,74
45		7,70	46		7,59
47		7,55	48		7,47

(Continua)

(Conclusão)

Tabela 11 – Estruturas e valores de pIC_{50} para o conjunto de dados 2.

Composto	Estrutura	pIC_{50}	Composto	Estrutura	pIC_{50}
49		7,30	50		7,19
51		7,11	52		6,62

Figuras geradas no programa ChemBioDraw 14.0.

Fonte: Elaborado pelo autor.

Análises PCA foram realizadas para o conjunto de dados 2 para avaliar a similaridade estrutural entre os compostos e a sua distribuição em relação aos valores de pIC_{50} . Como descrito para o conjunto 1, foi gerado um mapa de similaridade estrutural, a partir do qual foram selecionadas a série treinamento e a série teste. O mapa de similaridade estrutural para o conjunto 2 e a distribuição das moléculas de acordo com os valores de pIC_{50} , são mostrados na figura 17. O gráfico evidencia uma grande concentração de moléculas na faixa de pIC_{50} entre 7 e 8. O intervalo compreendido entre 8 e 9 é o que apresenta o menor número de moléculas. Os conjuntos treinamento e teste foram selecionados de modo a incluir em ambos moléculas de todas as regiões do mapa (representativas de todo o intervalo da propriedade alvo). Desta maneira foram selecionadas as seguintes moléculas para o conjunto teste: **4,13, 16, 17, 22, 27, 30, 36, 39, 46 e 49**.

Figura 17 – Mapa de similaridade estrutural para o conjunto 2 e distribuição das moléculas por faixa de atividade.

Fonte: Elaborada pelo autor.

4.2.2 Alinhamento Molecular

O alinhamento das moléculas do conjunto 2 foi realizado através do método Distill Rigid com base na subestrutura comum a todos os compostos do conjunto, utilizando como referência o composto mais potente (**31**). O alinhamento molecular para o conjunto 2 é ilustrado na figura 18.

Figura 18 – Alinhamento molecular para o conjunto de dados 2.

Fonte: Elaborada pelo autor.

4.2.3 Modelos de HQSAR – Conjunto 2

Modelos de HQSAR foram desenvolvidos para o conjunto de dados 2. Todos os comprimentos de holograma disponíveis (53 a 401) foram utilizados na construção dos modelos. Inicialmente, foram utilizadas as seguintes combinações dos parâmetros de distinção de fragmentos: A/B/C, A/C/DA, A/B/C/H/Ch, A/B/C/H/DA. Aplicando o tamanho de fragmento padrão de 4 a 7 átomos e o método de validação LOO, foram obtidos os resultados exibidos na tabela 12.

Tabela 12 – Resultados de HQSAR para o conjunto de dados 2, utilizando diferentes parâmetros de distinção de fragmentos, tamanho de fragmento de 4 a 7 átomos e método de validação LOO.

Distinção de Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
A/B/C	0,62	0,88	0,20	59	5
A/C/DA	0,50	0,84	0,23	307	4
A/B/C/H/Ch	0,66	0,96	0,14	61	13
A/B/C/H/DA	0,51	0,84	0,23	97	4

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *HL*, Comprimento do holograma; *N*, Número de componentes.

Fonte: Elaborada pelo autor.

Como pode ser observado, o melhor modelo foi obtido com a combinação A/B/C/H/Ch ($q^2 = 0,66$; $r^2 = 0,96$; e $SEE = 0,14$). O efeito da variação do tamanho dos fragmentos foi avaliado para o melhor modelo da tabela 12. A tabela 13 lista os resultados obtidos para os tamanhos de fragmento: 3–7, 4–7, 4–8, 5–9 e 6–9.

Tabela 13 – Resultados gerados para diferentes tamanhos de fragmento (modelo A/B/C/H/Ch).

Tamanho do Fragmento	Parâmetros Estatísticos					
	q^2	r^2	SEE	HL	N	r^2_{pred}
3-7	0,68	0,96	0,14	61	13	0,53
4-7	0,66	0,96	0,14	61	13	0,61
4-8	0,73	0,91	0,12	83	14	0,16
5-9	0,70	0,98	0,11	199	20	0,50
6-9	0,68	0,98	0,11	199	20	0,54

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; *SEE*, Erro padrão de estimativa; *HL*, Comprimento do holograma; *N*, Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

A variação do tamanho dos fragmentos não resultou em melhoria da consistência estatística. Para avaliar a estabilidade do melhor modelo, o conjunto de dados foi dividido aleatoriamente em cinco e dez grupos e a sua capacidade de predição foi avaliada pelo método LMO ($q^2 = 0,51$ e $0,66$ para cinco e dez grupos, respectivamente).

O processo de validação do melhor modelo de HQSAR (A/B/C/H/Ch, tabela 12) foi realizado por meio de uma série teste de compostos, selecionados a partir do mapa de similaridade estrutural (figura 17). O processo de validação foi aplicado a

este modelo para todos os tamanhos dos fragmentos listados na tabela 13. Os dados da tabela 13 demonstram que o modelo com maior capacidade preditiva é o que apresenta tamanho de fragmento de 4 a 7 átomos ($r^2_{\text{pred}} = 0,61$). A tabela 14 exhibe os valores experimentais e preditos para a série teste. Os resultados são mostrados graficamente na figura 19 para todo o conjunto de dados.

Tabela 14 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo de HQSAR - conjunto de dados 2).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
4	7,55	8,13	-0,58
13	6,60	6,73	-0,13
16	8,10	8,00	0,10
17	8,05	7,56	0,49
22	7,60	6,89	0,71
27	6,96	7,18	-0,22
30	6,06	6,33	-0,27
36	8,00	7,74	0,27
39	7,66	7,24	0,42
46	7,59	7,74	-0,15
49	7,30	6,98	0,32

^a Diferença entre os valores experimentais e preditos de pIC_{50} .

Fonte: Elaborada pelo autor.

Figura 19 – Valores experimentais *versus* preditos de pIC₅₀ para o melhor modelo de HQSAR (conjunto de dados 2).

Fonte: Elaborada pelo autor.

A boa correlação entre os valores experimentais e preditos de pIC₅₀ para o conjunto teste, dado pelo coeficiente de predição ($r^2_{\text{pred}} = 0,61$), demonstra a habilidade do melhor modelo em prever a atividade de novos compostos contra o *T. cruzi*.

Este modelo foi utilizado para avaliar a contribuição de diferentes fragmentos para a atividade dos compostos **2**, **8**, **14**, **15**, **24** e **45**. O grupo piperazina dos compostos foi destacado como um fragmento importante para a atividade biológica. O anel piridínico também foi assinalado como um grupo que contribui positivamente para a atividade biológica das moléculas do conjunto 2 (figura 20).

Figura 20 – Mapas de contribuição de HQSAR para alguns compostos do conjunto de dados 2.

Fonte: Elaborada pelo autor.

4.2.4 Modelos CoMFA – Conjunto 2

Modelos CoMFA foram desenvolvidos para investigar o modo pelo qual as características eletrostáticas e estereoquímicas determinam a atividade dos compostos do conjunto de dados 2. A consistência interna dos modelos foi avaliada com o método de validação cruzada LOO. Diversos modelos CoMFA foram gerados através do método SDC de focalização de região (tabela 15).

Tabela 15 – Modelos CoMFA obtidos através do método SDC de focalização de região (conjunto de dados 2).

SDC	q^2	SEP	r^2	SEE	N	r^2_{pred}
0,1	0,63	0,35	0,98	0,08	10	0,66
0,3	0,87	0,22	0,99	0,05	10	0,63
0,9	0,84	0,25	0,99	0,06	10	0,62

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEE, Erro padrão de estimativa; SEP, Erro padrão de predição; N, Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

Como pode ser observado, o modelo com maior capacidade preditiva para o conjunto teste ($r^2_{pred} = 0,66$) foi obtido ajustando-se os valores de SDC por um fator de 0,1. A consistência interna dos modelos também foi avaliada pelo método LMO ($q^2 = 0,56$ e $0,54$ para cinco e dez grupos, respectivamente). A diferença entre os valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA é exibida na tabela 16 e ilustrada graficamente na figura 21.

Tabela 16 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA - conjunto de dados 2).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
4	7,55	7,72	-0,17
13	6,60	7,09	-0,49
16	8,10	7,74	0,36
17	8,05	7,81	0,24
22	7,60	7,32	0,29
27	6,96	7,06	-0,10
30	6,06	6,79	-0,73
36	8,00	7,85	0,15

(Continua)

(Conclusão)

Tabela 16 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA - conjunto de dados 2).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
39	7,66	7,31	0,36
46	7,59	7,90	-0,30
49	7,3	7,00	0,30

^a Diferença entre os valores experimentais e preditos de pIC_{50} .

Fonte: Elaborada pelo autor.

Figura 21 – Valores experimentais *versus* preditos de pIC_{50} para o melhor modelo CoMFA (conjunto de dados 2).

Fonte: Elaborada pelo autor.

Como é possível observar na figura 21, o modelo CoMFA apresenta alta capacidade de correlação para o conjunto treinamento e de predição para o conjunto teste. Estes resultados confirmam sua utilidade para prever a atividade de novos compostos dentro da diversidade estrutural investigada.

A figura 22 mostra os mapas eletrostáticos e estereoquímicos para os compostos **2**, **14**, **18** e **45**. O mapa indica diferentes regiões onde a presença de grupos volumosos é desfavorável para a atividade (amarelo). Estas regiões desfavoráveis predominam sobre as regiões onde a presença de grupos volumosos está correlacionada com um aumento da atividade (verde). Em relação aos mapas

eletrostáticos, as regiões em que grupos eletronegativos são favoráveis (vermelho) predominam sobre aquelas em que grupos eletropositivos levam a um aumento na atividade dos compostos (azul). Observa-se que a região ao redor do anel piperazina concentra grande parte dos poliedros, principalmente os relacionados a grupos mais volumosos e eletronegativos. Este aspecto dos mapas de contorno indica a importância de grupos pouco volumosos e com carga parcial negativa como substituintes da piperazina.

Figura 22 – Mapas de contorno CoMFA para algumas moléculas do conjunto de dados 2.

Fonte: Elaborada pelo autor.

4.2.5 Modelos CoMSIA – Conjunto 2

Modelos CoMSIA foram desenvolvidos para o conjunto de dados 2. A consistência interna dos modelos foi avaliada pelo método LOO através da variação do fator de atenuação (tabela 17). O melhor modelo foi obtido aplicando-se um valor de 0,1 para o fator de atenuação.

Tabela 17 – Resultados CoMSIA obtidos através da variação do valor do fator de atenuação (conjunto de dados 2).

Fator de atenuação	q^2	SEP	r^2	SEE	N
0,1	0,41	0,46	0,90	0,19	8
0,2	0,31	0,47	0,86	0,22	5
0,3	0,29	0,48	0,88	0,20	5
0,4	0,28	0,49	0,90	0,18	5

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEE, Erro padrão de estimativa; SEP, Erro padrão de predição; N, Número de componentes.

Fonte: Elaborada pelo autor.

Como pode ser observado, esta abordagem resultou em valores de q^2 abaixo do esperado. Para eliminar os descritores redundantes e aumentar o peso dos que contribuem de maneira mais consistente para o modelo, foi aplicado o método de focalização de região. Os modelos gerados através deste procedimento são apresentados na tabela 18.

Tabela 18 – Resultados CoMSIA obtidos através da variação do valor de SDC (conjunto de dados 2).

SDC	q^2	SEP	r^2	SEE	N	r^2_{pred}
0,6	0,57	0,39	0,90	0,19	7	0,66
0,9	0,71	0,32	0,86	0,22	7	0,67

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEE, Erro padrão de estimativa; SEP, Erro padrão de predição; N, Número de componentes; r^2_{pred} , coeficiente de predição.

Fonte: Elaborada pelo autor.

Os dados da tabela 18 mostram que o modelo com valor de SDC = 0,9 é ligeiramente melhor em relação ao valor de r^2_{pred} . Portanto, a estabilidade deste modelo foi avaliada através do método LMO ($q^2 = 0,70$ e $0,74$ para cinco e dez grupos, respectivamente). Os valores experimentais e preditos de pIC_{50} para o

melhor modelo CoMSIA são listados na tabela 19 para o conjunto teste e ilustrados graficamente na figura 23 para todo o conjunto de dados 2.

Tabela 19 – Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA - conjunto de dados 2).

Composto	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
4	7,55	7,40	0,15
13	6,60	7,14	-0,54
16	8,10	8,11	-0,01
17	8,05	7,54	0,52
22	7,60	7,25	0,35
27	6,96	6,93	0,03
30	6,06	6,44	-0,38
36	8,00	8,02	-0,02
39	7,66	6,97	0,69
46	7,59	7,58	0,01
49	7,30	7,13	0,18

^a Diferença entre os valores experimentais e preditos de pIC_{50} .

Fonte: Elaborada pelo autor.

Figura 23 – Valores experimentais *versus* preditos de pIC_{50} para o melhor modelo CoMSIA (conjunto de dados 2).

Fonte: Elaborada pelo autor.

Os mapas de contorno indicam regiões em que alterações nas características eletrostáticas, estereoquímicas, hidrofóbicas, e a presença de grupos doadores e aceptores de interações de hidrogênio, determinam variações no valor da propriedade alvo. A figura 24 mostra os mapas CoMSIA para as moléculas **2**, **14**, **18** e **45**. O mapa azul indica a região em que a presença de grupos eletropositivos é favorável, ao passo que o poliedro amarelo indica que substituintes volumosos na posição *para* do anel piperazina são desfavoráveis para a atividade biológica. Estes resultados são compatíveis com os mapas CoMFA (figura 22).

Figura 24 – Mapas de contorno estereoquímicos e eletrostáticos para o melhor modelo CoMSIA (conjunto de dados 2).

Fonte: Elaborada pelo autor.

Na figura 25 são identificadas as características hidrofóbicas e a presença de átomos doadores e aceptores de interações de hidrogênio. Os poliedros brancos indicam que grupos hidrofóbicos na região do anel piridínico e substituintes na posição *para* da piperazina, são desfavoráveis para a atividade. De acordo com o mapa, grupos aceptores de interações de hidrogênio são desfavoráveis como

substituintes da piperazina, ao passo que grupos doadores nesta região, estão correlacionados ao aumento da atividade.

Figura 25 – Mapas de contorno para doadores e receptores de interações de hidrogênio, e para grupos hidrofóbicos, gerados pelo modelo CoMSIA, para algumas moléculas do conjunto de dados 2.

Fonte: Elaborada pelo autor.

Conclusões e Perspectivas

5 Conclusões e Perspectivas

Nesta dissertação de mestrado foram investigadas as relações quantitativas entre a estrutura e atividade de dois conjuntos de compostos com atividade *in vitro* contra o protozoário *Trypanosoma cruzi*. Os compostos investigados são inspirados na estrutura do fenarimol, e constituem uma classe química promissora para o desenvolvimento de novos agentes para o tratamento da doença de Chagas. Foram utilizados os métodos de QSAR 2D HQSAR e os métodos de QSAR 3D CoMFA e CoMSIA. Estas estratégias proporcionaram a compreensão de aspectos moleculares essenciais, responsáveis pela atividade destas séries de compostos. Dentre os resultados obtidos, os seguintes aspectos se destacam:

- ✓ Os modelos de QSAR apresentam alta consistência interna e externa, como indicado pelos coeficientes de correlação.
- ✓ O emprego dos conjuntos teste revelou a habilidade dos modelos em prever a atividade de novos compostos estruturalmente relacionados ao fenarimol.
- ✓ Os mapas de contribuição 2D, gerados pelo modelo de HQSAR, identificaram fragmentos moleculares importantes para a atividade dos compostos investigados
- ✓ Os mapas de contorno 3D gerados, pelos modelos CoMFA e CoMSIA, evidenciaram características relevantes para a atividade dos conjuntos investigados. Foram revelados aspectos eletrostáticos, estereoquímicos, hidrofóbicos, e relacionados a presença de doadores e aceptores de interações de hidrogênio.
- ✓ As características assinaladas pelos mapas 2D e 3D são recursos úteis para delimitar os espaço químico a ser explorado no planejamento de novos compostos dentro da diversidade estrutural estudada.
- ✓ Os modelos de QSAR desenvolvidos nesse trabalho poderão ser utilizados para guiar futuros estudos de química medicinal do grupo, na síntese e avaliação biológica de novos candidatos a fármacos para o tratamento da

doença de Chagas. Novos compostos serão planejados de acordo com as modificações moleculares destacadas na figura 26.

Figura 26 – Estratégias a serem exploradas em trabalhos de otimização molecular para esta série de compostos. A) Composto **41** do conjunto de dados 1. B) Composto **18** do conjunto de dados 2.

Fonte: Elaborado pelo autor.

Referências

REFERÊNCIAS

- 1 WERMUTH, C. G. **The practice of medicinal chemistry**. 3rd. ed. London: Academic Press, 2008. p.63.
- 2 SURADE, S.; BLUNDELL, T. L. Structural biology and drug discovery of difficult targets: the limits of ligandability. **Chemistry & Biology**, v. 19, n. 1, p. 42-50, 2012.
- 3 DROR, R. O. et al. Biomolecular simulation: a computational microscope for molecular biology. **Annual Review of Biophysics**, v. 41, p. 429-452, 2012.
- 4 ANDRICOPULO, A. D.; SALUM, L. B.; ABRAHAM, D. J. Structure-based drug design strategies in medicinal chemistry. **Current Topics in Medicinal Chemistry**, v. 9, n. 9, p. 771-790, 2009.
- 5 ACHARYA, C. et al. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. **Current Computer-Aided Drug Design**, v. 7, n. 1, p. 10, 2011.
- 6 KALYAANAMOORTHY, S.; CHEN, Y. P. Phoebe. Structure-based drug design to augment hit discovery. **Drug Discovery Today**, v. 16, n. 17, p. 831-839, 2011.
- 7 YANG, S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug Discovery Today**, v. 15, n. 11, p. 444-450, 2010.
- 8 FERREIRA, L. L. G. **Identificação de novos inibidores da enzima aldolase de Trypanosoma brucei**. 2013. 164 p. Tese (Doutorado em Física Aplicada) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- 9 MUNTEANU, C. R. et al. Drug discovery and design for complex diseases through QSAR computational methods. **Current Pharmaceutical Design**, v. 16, n. 24, p. 2640-2655, 2010.
- 10 VILLOUTREIX, B. O.; EUDES, R.; MITEVA, M. A. Structure-based virtual ligand screening: recent success stories. **Combinatorial Chemistry & High Throughput Screening**, v. 12, n. 10, p. 1000-1016, 2009.
- 11 CHENG, T. et al. Structure-based virtual screening for drug discovery: a problem-centric review. **AAPS Journal**, v. 14, n. 1, p. 133-141, 2012.
- 12 GASSMANN, O.; ENKEL, E.; CHESBROUGH, H. The future of open innovation. **R&D Management**, v. 40, n. 3, p. 213-221, 2010.
- 13 DICKSON, M.; GAGNON, J. P. The cost of new drug discovery and development. **Discovery Medicine**, v. 4, n. 22, p. 172-179, 2009.

- 14 PAMMOLLI, F.; MAGAZZINI, L.; RICCABONI, M. The productivity crisis in pharmaceutical R&D. **Nature Reviews Drug Discovery**, v. 10, n. 6, p. 428-438, 2011.
15. PUZYN, T.; LESZCZYNSKI, J.; CRONIN, M.T. **Recent advances in QSAR studies**: methods and applications. Dordrecht: Springer, 2010.p. 3-4. (Challenges and advances in computational chemistry and physics, v. 8). doi: 10.1007/978-1-4020-9783-6
- 16 HONORIO, K. M.; MODA, T. L.; ANDRICOPULO, A. D. Pharmacokinetic properties and in silico Adme modeling In drug discovery. **Medicinal Chemistry**, v. 9, n. 2, p. 163-176, 2013.
- 17 MUNTEANU, C. R. et al. Drug discovery and design for complex diseases through QSAR computational methods. **Current Pharmaceutical Design**, v. 16, n. 24, p. 2640-2655, 2010.
- 18 KHAN, M. T. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. **Current Drug Metabolism**, v. 11, n. 4, p. 285-295, 2010.
- 19 SEDYKH, A. et al. Use of in vitro HTS-derived concentration a response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. **Environmental Health Perspectives**, v. 119, n. 3, p. 364-370, 2010.
- 20 ISTYASTONO, E. P. et al. Molecular determinants of ligand binding modes in the histamine H4 receptor: linking ligand-based three-dimensional quantitative structure–activity relationship (3D-QSAR) models to in silico guided receptor mutagenesis studies. **Journal of Medicinal Chemistry**, v. 54, n. 23, p. 8136-8147, 2011.
- 21 LEWIS, R. A.; WOOD, D. Modern 2D QSAR for drug discovery. **Wiley Interdisciplinary Reviews: computational molecular science**, v. 4, n. 6, p. 505-522, 2014. doi: 10.1002/wcms.1187.
- 22 TEFFEN, A. et al. Comparison of molecular fingerprint methods on the basis of biological profile data. **Journal of Chemical Information and Modeling**, v. 49, n. 2, p. 338-347, 2009.
- 23 POLANSKI, J. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. **Current Medicinal Chemistry**, v. 16, n. 25, p. 3243-3257, 2009.
- 24 TROPSHA, A. Best practices for QSAR model development, validation, and exploitation. **Molecular Informatics**, v. 29, n. 6-7, p. 476-488, 2010.
- 25 KIRALJ, R.; FERREIRA, M. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of the Brazilian Chemical Society**, v. 20, n. 4, p. 770-787, 2009.

- 26 HQSAR™. Disponível em: <http://www.tripos.com/data/SYBYL/HQSAR_072505.pdf>. Acesso em: 12 dez. 2014.
- 27 SALUM, L. B.; ANDRICOPULO, A. D. Fragment-based QSAR: perspectives in drug design. **Molecular Diversity**, v. 13, n. 3, p. 277-285, 2009.
- 28 ABDI, H. Partial least squares regression and projection on latent structure regression (PLS Regression). **Wiley Interdisciplinary Reviews: computational statistics**, v. 2, n. 1, p. 97-106, 2010.
- 29 SANTOS CRUZ, D.; SANTOS CASTILHO, M. 2D QSAR studies on series of human Beta-secretase (BACE-1) inhibitors. **Medicinal Chemistry**, v. 10, n. 2, p. 162-173, 2014.
- 30 VERMA, J.; KHEDKAR, V. M.; COUTINHO, E. C. 3D-QSAR in drug design-a review. **Current Topics in Medicinal Chemistry**, v. 10, n. 1, p. 95-115, 2010.
- 31 MITTAL, R. R. et al. Partial charge calculation method affects CoMFA QSAR prediction accuracy. **Journal of Chemical Information and Modeling**, v. 49, n. 3, p. 704-709, 2009.
- 32 ZHANG, L. et al. How to generate reliable and predictive CoMFA models. **Current Medicinal Chemistry**, v. 18, n. 6, p. 923-930, 2011.
- 33 MITTAL, R. R.; MCKINNON, R. A.; SORICH, M. J. The effect of molecular fields, lattice spacing and analysis options on CoMFA predictive ability. **QSAR & Combinatorial Science**, v. 28, n. 6-7, p. 637-644, 2009.
- 34 CHAUDHAERY, S. S.; ROY, K. K.; SAXENA, A. K. Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors. **Journal of Chemical Information and Modeling**, v. 49, n. 6, p. 1590-1601, 2009.
- 35 WOLBER, G.; DORNHOFER, A. A.; LANGER, T. Efficient overlay of small organic molecules using 3D pharmacophores. **Journal of Computer-Aided Molecular Design**, v. 20, n. 12, p. 773-788, 2006.
- 36 APARNA, V. et al. 3D-QSAR studies on antitubercular thymidine monophosphate kinase inhibitors based on different alignment methods. **Bioorganic & Medicinal Chemistry Letters**, v. 16, n. 4, p. 1014-1020, 2006.
- 37 KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. **Journal of Medicinal Chemistry**, v. 37, n. 24, p. 4130-4146, 1994.
- 38 FEASEY, N. et al. Neglected tropical diseases. **British Medical Bulletin**, v. 93, n. 1, p. 179-200, 2010.

- 39 CONTEH, L.; ENGELS, T.; MOLYNEUX, D. H. Socioeconomic aspects of neglected tropical diseases. **Lancet**, v. 375, n. 9710, p. 239-247, 2010.
- 40 CROMPTON, D. WT. **Sustaining the drive to overcome the global impact of neglected tropical diseases**: second WHO report on neglected tropical diseases. Geneva: WHO, 2013.
- 41 HOTEZ, P. J. et al. Chagas disease: "the new HIV/AIDS of the Americas". **PLoS Neglected Tropical Diseases**, v. 6, n. 5, p. e1498, 2012.
- 42 RASSI JUNIOR, A.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **Lancet**, v. 375, n. 9723, p. 1388-1402, 2010.
- 43 DIAS, J. C. P. The treatment of Chagas disease (South American trypanosomiasis). **Annals of Internal Medicine**, v. 144, n. 10, p. 772-774, 2006.
- 44 CASTRO, J. A.; MONTALTO DEMECCA, M.; BARTEL, L. C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). **Human & Experimental Toxicology**, v. 25, n. 8, p. 471-479, 2006.
- 45 DNDi. **Chagas diseases**. Disponível em: <<http://www.dndi.org/diseases-projects/diseases/chagas/target-product-profile.html>> Acesso em: 20 dez. 2014
- 46 CHATELAIN, E.; IOSET, J-R. Drug discovery and development for neglected diseases: the DNDi model. **Drug Design, Development and Therapy**, v. 5, p. 175, 2011. doi: 10.2147/DDDT.S16381.
- 47 BRUN, R. et al. Human african trypanosomiasis. **Lancet**, v. 375, n. 9709, p. 148-159, 2010.
- 48 SUNDAR, S.; CHAKRAVARTY, J. Leishmaniasis: an update of current pharmacotherapy. **Expert Opinion on Pharmacotherapy**, v. 14, n. 1, p. 53-63, 2013.
- 49 PLANER, J. D. et al. Synergy testing of FDA-approved drugs identifies potent drug combinations against *Trypanosoma cruzi*. **PLoS Neglected Tropical Diseases**, v. 8, n. 7, p. e2977, 2014.
- 50 MAERTENS, J. A. History of the development of azole derivatives. **Clinical Microbiology and Infection**, v. 10, n. s1, p. 1-10, 2004.
- 51 DOCAMPO, R. et al. Biochemical and ultrastructural alterations produced by miconazole and econazole in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 3, n. 3, p. 169-180, 1981.
- 52 URBINA, J. A. Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. **Acta Tropical**, v. 115, n. 1, p. 55-68, 2010.
- 53 NOSTEN, F. et al. Antimalarial drugs in pregnancy: a review. **Current Drug Safety**, v. 1, n. 1, p. 1-15, 2006.

- 54 WITSCHER, M. et al. Agrochemicals against malaria, sleeping sickness, leishmaniasis and Chagas disease. **PLoS Neglected Tropical Diseases**, v. 6, n. 10, p. e1805, 2012.
- 55 KEENAN, M. et al. Analogues of fenarimol are potent inhibitors of *Trypanosoma cruzi* and are efficacious in a murine model of Chagas disease. **Journal of Medicinal Chemistry**, v. 55, n. 9, p. 4189-4204, 2012.
- 56 KEENAN, M. et al. Two analogues of fenarimol show curative activity in an experimental model of Chagas disease. **Journal of Medicinal Chemistry**, v. 53, n. 24, p. 10158-10170, 2013.
- 57 HARGROVE, T. Y. et al. Complexes of *Trypanosoma cruzi* sterol 14 α -demethylase (cyp51) with two pyridine-based drug candidates for Chagas disease. **Journal of Biological Chemistry**, v. 288, n. 44, p. 31602-31615, 2013.
- 58 MODA, T. L.; MONTANARI, C. A.; ANDRICOPULO, A. D. Hologram QSAR model for the prediction of human oral bioavailability. **Bioorganic & Medicinal Chemistry**, v. 15, n. 24, p. 7738-7745, 2007.
- 59 TROSSINI, G. H. G. et al. Quantitative structure–activity relationships for a series of inhibitors of cruzain from *Trypanosoma cruzi*: molecular modeling, CoMFA and CoMSIA studies. **Journal of Molecular Graphics and Modelling**, v. 28, n. 1, p. 3-11, 2009.
- 60 FERREIRA, L. G.; ANDRICOPULO, A. D. Structure-and ligand-based structure-activity relationships for a series of inhibitors of aldolase. **Current Computer-Aided Drug Design**, v. 8, n. 4, p. 309-316, 2012.
- 61 KEENAN, M. et al. Design, structure–activity relationship and in vivo efficacy of piperazine analogues of fenarimol as inhibitors of *Trypanosoma cruzi*. **Bioorganic & Medicinal Chemistry**, v. 21, n. 7, p. 1756-1763, 2013.
- 62 KEENAN, M. et al. Two analogues of fenarimol show curative activity in an experimental model of Chagas disease. **Journal of Medicinal Chemistry**, v. 56, n. 24, p. 10158-10170, 2013.
- 63 POWELL, M. J. D. Restart procedures for conjugate gradient method. **Mathematical Programming**, v.12, p.241-254,1977.
- 64 TSAI, K. et al. A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. **European Journal of Medicinal Chemistry**, v. 45, n. 4, p. 1544-1551, 2010.
- 65 CLARK, M.; CRAMER III, R. D.; VAN OPDENBOSCH, N. Validation of the general purpose tripos 5.2 force field. **Journal of Computational Chemistry**, v.10, p.982-1012, 1989.

66 TRIPOS Go. Discover. Distill: Determine and visualize structure-activity relationship. 2005. Disponível em: <<http://www.tri-ibiotech.com.cn/upload/file/sybyl-x/9-Distill.pdf>> Acesso em: 23 dez. 2014.

67 ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: computational statistics**, v. 2, n. 4, p. 433-459, 2010.

68 TRIPOS Go. Discover. UNITY: Locate compounds in databases that match a pharmacophore or fit receptor site. 2008. Disponível em : <http://www.tripos.com/tripos_resources/fileroot/pdfs/Unity_111408.pdf>. Acesso em: 23 dez. 2014.

69 TRIPOS Go. Discover . Advanced CoMFA: Refine and enhance 3D QSAR models. 2005. Disponível em: <http://tripos.com/data/SYBYL/AdvancedComfa_072505.pdf>. Acesso em: 24 dez. 2014.

70 HAWKINS, D. M.; BASAK, S. C.; MILLS, D. Assessing model fit by cross-validation. **Journal of Chemical Information and Computer Sciences**, v. 43, n. 2, p. 579-586, 2003.