UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

CÉSAR MASCHIO FIORAVANTI

Estudos de modelagem molecular para uma série de derivados
piridínicos com ação antichagásica

São Carlos
2015

CÉSAR MASCHIO FIORAVANTI

Estudos de modelagem molecular para uma série de derivados
piridínicos com ação antichagásica

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)

Fioravanti, César Maschio

Estudos de modelagem molecular para uma série de derivados piridínicos com ação antichagásica / César Maschio Fioravanti; 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. Doença de Chagas. 2. Modelagem molecular. 3. QSAR. 4. Química medicinal. I. Andricopulo, Adriano Defini, orient. II. Título.

Aos meus pais Carlos e Regina e à minha irmã Lívia.

AGRADECIMENTOS

Aos meus pais, Carlos Henrique Fioravanti e Regina Maura Maschio Fioravanti, e à minha irmã, Lívia Maschio Fioravanti.

Ao Prof. Dr. Adriano D. Andricopulo pela orientação, oportunidade e confiança.

Ao Prof. Dr. Rafael V. C. Guido pelo auxílio no desenvolvimento deste trabalho.

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

Aos colegas do Laboratório de Química Medicinal e Computacional, Ricardo N. dos Santos, Ivani Pauli, Mariana Laureano, Karina Matos, Anacleto Souza, Luma Godoy Magalhães, Renata Krogh, Simone M. Michelin, Vanessa Kiraly, Andrew Oliveira, Fernando V. Maluf, Renata Bueno e Gustavo A. Lima.

Aos meus amigos e companheiros de república, Alexandre Adele, Bruno Nodomi, Bruno Takeyama, Danilo Zamprônio, Erik Hissanori, Felipe Kodama, Fernando Tateno, Guilherme Hashimoto, Rafael Miyamoto, Ricardo Kojima e Rodrigo Hashizume.

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

Ao Prof. Luiz Carlos Dias e ao Dr. Marco Aurélio Dessoay (IQ-UNICAMP), pela síntese, e a Dra. Renata Krogh (LQMC-IFSC), pela avaliação biológica, dos compostos cujos dados estrutura-atividade foram empregados nesta dissertação para geração dos modelos de QSAR 2D e 3D.

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

RESUMO

FIORAVANTI, C. M. **Estudos de modelagem molecular para uma série de derivados piridínicos com ação antichagásica**. 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 doença tropical negligenciada de alta prioridade nos programas de pesquisa e desenvolvimento da Organização Mundial da Saúde (OMS). Endêmica em 21 países da América Latina, a doença tem recentemente se tornado uma preocupação de saúde em outras regiões do planeta, como a América do Norte, Europa, Oceania e Ásia. Causada pelo parasita *Trypanosoma cruzi*, a doença de Chagas atinge 8-10 milhões de pessoas, de acordo com estimativas da OMS. Os únicos dois medicamentos disponíveis para o tratamento da doença são ineficazes e causam graves efeitos adversos. Este panorama tem evidenciado a urgência por novos agentes terapêuticos. O objetivo desta dissertação de mestrado é o desenvolvimento de modelos de relações quantitativas entre a estrutura e atividade (QSAR, na sigla inglesa para *quantitative structure-activity relationships*) para uma série de derivados piridínicos inspirados em um composto líder com potente ação anti-*T. cruzi*. Os modelos de QSAR foram gerados com os métodos holograma QSAR (HQSAR, na sigla inglesa para *hologram QSAR*); análise comparativa dos campos moleculares (CoMFA, na sigla inglesa *comparative molecular field analysis*); e análise comparativa dos índices de similaridade estrutural (CoMSIA, na sigla inglesa para *comparative molecular similarity indices analysis*). Os modelos identificaram as principais características estruturais que determinam a potência biológica destes compostos. Além disso, os modelos possuem alta consistência estatística e elevada capacidade de predição para novas moléculas. Os resultados apresentados revelam a utilidade dos modelos de QSAR no planejamento de novos antichagásicos mais potentes dentro da diversidade química explorada.

Palavras-chave: Doença de Chagas. Modelagem molecular. QSAR. Química medicinal.

ABSTRACT

FIORAVANTI, C. M. **Molecular modeling studies for a series of pyridine derivatives with antichagasic action.** 2015. 104 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

Chagas' disease is a neglected tropical disease of high priority for the World Health Organization (WHO) drug discovery and development programs. Endemic in 21 countries in Latin America, the disease has recently become a public health problem in other regions of the planet, such as North America, Europe, Oceania and Asia. Caused by the parasite *Trypanosoma cruzi*, Chagas' disease affects 8-10 million people worldwide, according to WHO. The only two available drugs are ineffective and cause serious side effects. These drawbacks have demonstrated the urgency for novel therapeutic agents. The main goal of this master's dissertation is the development of quantitative structure-activity relationships models (QSAR) for a series of pyridine derivatives having prominent activity against *T. cruzi*. The compounds were inspired in the structure of a lead compound with potent anti-*T. cruzi* action. The QSAR models were created using the methods hologram QSAR (HQSAR); comparative molecular field analysis (CoMFA); and comparative molecular similarity indices analysis (CoMSIA). The resulting models have identified key molecular features that determine the biological potency of the compounds. Furthermore, these models possess high statistical consistency and distinct ability to predict the potency of new molecules. The results reported reveal the usefulness of the developed QSAR models in the design of new antichagasic agents within this structural diversity.

Keywords: Chagas' disease. Molecular modeling. QSAR. Medicinal chemistry.

LISTA DE FIGURAS

Figura 1 -	Distribuição global da doença de Chagas.	24
Figura 2 -	Fármacos utilizados no tratamento da doença de Chagas: nifurtimox (1) e benzonidazol (2).	25
Figura 3 -	Esquematisação do ciclo de vida do <i>T. cruzi</i> .	27
Figura 4 -	Etapas da construção de um modelo de HQSAR.	33
Figura 5 -	Esquema de construção de modelos CoMFA e CoMSIA. As moléculas do conjunto de dados são alocadas em uma grade tridimensional e uma sonda é utilizada como átomo de prova.	35
Figura 6 -	Farmacóforo característico da série de compostos utilizada na modelagem de QSAR.	51
Figura 7 -	Distribuição dos valores da propriedade biológica nos conjuntos teste, treinamento e no conjunto total.	60
Figura 8 -	Valores experimentais e valores preditos de pIC_{50} para o melhor modelo de HQSAR.	65
Figura 9 -	Mapas de contribuição de HQSAR para os compostos 88, 101, 108 e 89.	66
Figura 10 -	Mapas de contribuição de HQSAR para os compostos 78,79, 80 e 96.	67
Figura 11 -	Alinhamento molecular obtido pelo método MCS.	68
Figura 12 -	Compostos utilizados na construção do modelo farmacofórico.	69
Figura 13 -	Alinhamento gerado com o GALAHAD. As esferas dos anéis, em verde claro, indicam os grupos aromáticos, e as outras duas, em verde, indicam aceptores de interações de hidrogênio.	70
Figura 14 -	Alinhamento por farmacóforos.	70
Figura 15 -	Alinhamento molecular utilizando estruturas de mínima energia.	72
Figura 16-	Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 1).	76

Figura 17 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 2).	79
Figura 18 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 3).	82
Figura 19 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMSIA (alinhamento 1).	86
Figura 20 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMSIA (alinhamento 2).	89
Figura 21 - Mapas de contorno CoMFA para o composto 85. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2 e (C) Alinhamento 3.	91
Figura 22 - Estrutura e atividade dos compostos 85, 88, e 91, ilustrando a diminuição gradativa do volume dos grupos na região do anel aromático. Essa diminuição é acompanhada por uma redução nos valores de atividade.	91
Figura 23 - Mapas de contorno CoMSIA eletrostáticos e estereoquímicos para o composto 85. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2.	92
Figura 24 - Mapas de contorno CoMSIA hidrofóbicos, doares e aceptores de interações de hidrogênio para o composto 85. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2.	92
Figura 25 - Estratégias a serem exploradas em trabalhos de otimização molecular para esta série de compostos.	98

LISTA DE TABELAS

Tabela 1 -	Principais parâmetros estatísticos utilizados para se avaliar a qualidade de um modelo de QSAR.	31
Tabela 2 -	Parâmetros de distinção de fragmentos no método HQSAR.	33
Tabela 3 -	Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .	52
Tabela 4 -	Resultados de HQSAR para diferentes parâmetros de distinção de fragmentos, utilizando o tamanho dos fragmentos de 7 a 10 átomos.	61
Tabela 5 -	Modelos estatísticos gerados pelo método de HQSAR.	62
Tabela 6 -	Resultados de HQSAR para diferentes tamanhos dos fragmentos (modelo A/C/DA).	63
Tabela 7 -	Resultados LMO para o modelo A/C/DA com tamanho dos fragmentos de 7 a 10 átomos.	63
Tabela 8 -	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo de HQSAR).	64
Tabela 9 -	Resultados CoMFA obtidos com o método de focalização de região SDC.	73
Tabela 10 -	Resultados CoMFA em função da variação do número de componentes (SDC = 0,7; alinhamento 1).	74
Tabela 11 -	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 1).	74
Tabela 12 -	Resultados CoMFA obtidos em função da variação do número de componentes (SDC = 0,7; alinhamento 2).	77
Tabela 13 -	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 2).	77
Tabela 14 -	Resultados CoMFA obtidos em função da variação do número de componentes (SDC = 0,7; alinhamento 3).	80
Tabela 15 -	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 3).	80

Tabela 16 -	Resultados CoMSIA obtidos através do método SDC.	83
Tabela 17 -	Resultados CoMSIA obtidos em função do número de componentes (SDC = 0,5; alinhamento 1).	84
Tabela 18-	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA – alinhamento 1).	85
Tabela 19 -	Resultados CoMSIA obtidos em função da variação do número de componentes (SDC = 0,5; alinhamento 2).	87
Tabela 20 -	Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA – alinhamento 2).	87

LISTA DE ABREVIATURAS E SIGLAS

DTNs	Doenças tropicais negligenciadas
OMS	Organização Mundial da Saúde
P&D	Pesquisa e desenvolvimento
PPPs	Parcerias público-privadas
TDR	<i>Special Programme for Research and Training in Tropical Diseases)</i>
DNDi	<i>Drugs for Neglected Diseases Initiative</i>
ADME	Absorção, distribuição, metabolismo e excreção
NCE	<i>New Chemical Entities</i>
SBDD	<i>Structure-based Drug Design</i>
VS	<i>Virtual Screening</i>
LBDD	<i>Ligand-based Drug Design</i>
QSAR	<i>Quantitative Structure-Activity Relationships</i>
r^2	Coeficiente de correlação
q^2	Correlação com validação cruzada
r^2_{pred}	Coeficiente de correlação preditivo
LOO	<i>Leave-One-Out</i>
LMO	<i>Leave-Many-Out</i>
HQSAR	<i>Hologram Quantitative Structure-Activity Relationships</i>
A	Átomos
B	Ligações
C	Conectividade
H	Hidrogênios
Ch	Quiralidade
DA	Doador e Aceptor
PLS	<i>Partial Least Squares</i>
CoMFA	<i>Comparative Molecular Field Analysis</i>
CoMSIA	<i>Comparative Molecular Similarity Indices Analysis</i>
GALAHAD	<i>Genetic Algorithm with Linear Assignment for Hypermolecular Alignment of Datasets</i>

GASP	<i>Genetic Algorithm Similarity Program</i>
2D	Bidimensional
3D	Tridimensional
IC ₅₀	Concentração necessária para inibir 50% a reprodução do parasita em cultura
SDC	<i>Standard deviation coefficient</i>
SEP	Erro padrão de predição
SEE	Erro padrão de estimativa
HL	Comprimento do holograma
N	Número de componentes
MCS	<i>Maximum Common Substructure</i>

SUMÁRIO

1	Introdução	21
1.1	Doença Tropicais Negligenciadas	23
1.2	Doença de Chagas	24
1.2.1	<i>Ciclo de Transmissão</i>	25
1.2.2	<i>Aspectos Clínicos da Doença de Chagas</i>	27
1.3	Descoberta de Fármacos	28
1.4	Modelagem Molecular e Planejamento de Fármacos	28
1.5	Relações Quantitativas entre Estrutura e Atividade	29
1.5.1	<i>Validação dos modelos de QSAR</i>	30
1.5.2	<i>QSAR Bidimensional - HQSAR</i>	32
1.5.3	<i>QSAR Tridimensional – CoMFA e CoMSIA</i>	34
1.5.4	<i>Alinhamento Estrutural</i>	36
2	Objetivos	37
3	Materiais e Métodos	41
3.1	O Conjunto de Dados	43
3.2	Estruturas Tridimensionais	44
3.3	Alinhamento Molecular	44
3.4	HQSAR	45
3.5	CoMFA	46
3.6	CoMSIA	47
3.7	Validação Interna e Externa dos Modelos	47

4 Resultados e Discussão	49
4.1 O Conjunto de Dados	51
4.2 HQSAR	61
4.2.1 Mapas de Contribuição	65
4.3 Alinhamento Molecular	67
4.3.1 Alinhamento baseado na máxima subestrutura comum (Alinhamento 1)	67
4.3.2 Alinhamento baseado na superposição de farmacóforos (Alinhamento 2)	68
4.3.3 Alinhamento baseado nas estruturas de mínima energia (Alinhamento 3)	71
4.4 CoMFA	72
4.4.1 Modelos CoMFA – Alinhamento 1	72
4.4.2 Modelos CoMFA – Alinhamento 2	76
4.4.3 Modelos CoMFA – Alinhamento 3	79
4.5 CoMSIA	83
4.5.1 Modelos CoMSIA – Alinhamento 1	83
4.5.2 Modelos CoMSIA – Alinhamento 2	86
4.6 Mapas de Contorno	90
5 Conclusões e Perspectivas	95
Referências	99

Introdução

1 Introdução

1.1 Doenças Tropicais Negligenciadas

As doenças tropicais negligenciadas (DTNs) são consideradas um grande problema de saúde pública mundial. Segundo dados da Organização Mundial da Saúde (OMS), as DTNs afetam mais de 1 bilhão de pessoas em todo o planeta, com ampla predominância em países em desenvolvimento com climas tropicais e subtropicais.^{1,2} Zonas rurais remotas, áreas urbanas de baixa renda e regiões de conflito militar são os principais nichos de concentração das DTNs. Doenças como a esquistossomose, dengue, leishmaniose e doença de Chagas, fazem parte do grupo das 17 DTNs, de acordo com a classificação da OMS.^{1,3}

O descaso global com este grupo de doenças pode ser observado nos baixos níveis de investimentos destinados a pesquisa e desenvolvimento (P&D) de novos fármacos. Apenas uma pequena parcela dos recursos de P&D das companhias farmacêuticas é investida em DTNs. O baixo poder aquisitivo dos países fortemente afetados por estas doenças e a consequente falta de perspectivas de retorno financeiro dos altos investimentos no processo de descoberta de fármacos, são os principais obstáculos para o desenvolvimento de pesquisas na área.

Os prejuízos econômicos causados por este grupo de doenças não se restringem aos gastos com prevenção e tratamento. A perda de força de trabalho e o comprometimento em produtividade são responsáveis por grandes perdas econômicas nas regiões endêmicas.⁴ Deste modo, as DTNs são reconhecidamente um grande entrave para o desenvolvimento socioeconômico dos países atingidos.²

Na tentativa de reverter esse cenário preocupante, as parcerias público-privadas são uma modalidade atrativa de fomento a pesquisa na área de DTNs. Iniciativas como o Programa Especial de Treinamento e Pesquisa em Doenças Negligenciadas (TDR, na sigla inglesa para *Special Programme for Research and Training in Tropical Diseases*) da OMS, e a Iniciativa Fármacos para Doenças Negligenciadas (DNDi, na sigla inglesa para *Drugs for Neglected Diseases Initiative*),

entre outras, são exemplos importantes de ações internacionais que englobam parcerias multi-institucionais e interdisciplinares.⁵

1.2 Doença de Chagas

A doença de Chagas é uma infecção causada pelo protozoário *Trypanosoma cruzi*, que afeta cerca de 8 milhões de pessoas em todo o planeta.² Endêmica em 21 países da América Latina, incluindo Brasil, Argentina, Uruguai e México, a doença também tem sido encontrada em outras regiões do mundo. Devido aos crescentes índices de migração, a doença de Chagas é, atualmente, um problema de saúde em países da América do Norte, Europa e Ásia.⁶ A figura 1 mostra a distribuição global da doença, bem como o número estimado de casos nas diferentes regiões em que se tem registro.

Figura 1- Distribuição global da doença de Chagas.

Fonte: Adaptado de WHO, 2013.²

Apesar da gravidade do panorama atual e dos mais de 100 anos de sua descoberta, ainda não há qualquer tratamento adequado para a doença de Chagas.^{7,8} Os únicos dois medicamentos disponíveis, nifurtimox (Lampit[®], Bayer) e benzonidazol (Rochagan[®], Roche), foram desenvolvidos na década de 1970 e apresentam problemas como baixa eficácia e sérios efeitos colaterais. Ademais, estes compostos não são eficazes em estágios avançados da doença (fase crônica).^{9,10} A figura 2 exibe a estrutura destes dois fármacos.

Figura 2- Fármacos utilizados no tratamento da doença de Chagas: nifurtimox (1) e benzonidazol (2).

Fonte: Elaborado pelo autor.

1.2.1 Ciclo de transmissão

O principal mecanismo de transmissão da doença de Chagas é através da picada de insetos vetores, que estão infectados com o parasita.⁹ Além disso, a doença pode ser transmitida por transfusões de sangue contaminado e da mãe para o filho pela placenta. Casos de transmissão por via oral também foram registrados por meio da ingestão de alimentos contaminados pelo parasita, como o caldo de cana.

A transmissão vetorial ocorre através de insetos triatomíneos, popularmente conhecidos como “barbeiros”, que pertencem à família *Reduviidae*. Mais de 100 espécies de triatomíneos transmitem o *T. cruzi*, mas apenas algumas habitam locais com presença humana significativa.¹¹ Dentre elas estão o *Triatoma infestans*, *Triatoma dimidiata* e *Rhodnius prolixus*.¹² Estas espécies são comuns em regiões

com casas de barro e de pau-a-pique, usualmente encontradas em áreas rurais e centros urbanos empobrecidos.¹¹

Em seu complexo ciclo de vida, o *T. cruzi* assume três formas distintas: tripomastigota, amastigota e epimastigota. A forma infectante tripomastigota é capaz de penetrar nas células e tecidos. O amastigota é responsável pela proliferação no interior das células infectadas, enquanto que o epimastigota está presente no intestino delgado do inseto vetor. Ao picar um indivíduo portador da doença de Chagas, o triatomíneo ingere sangue contendo tripomastigotas. O parasita migra para o intestino delgado do inseto, onde se diferencia em epimastigota e se reproduz rapidamente. Os epimastigotas migram para o intestino grosso e se diferenciam em tripomastigotas metacíclicos. Ao se alimentar de sangue, o barbeiro defeca próximo ao local da picada liberando tripomastigotas. A seguir, o indivíduo se coça e o *T. cruzi* penetra no local da picada, ou ainda, nas regiões dos olhos e boca.¹¹ Após a infecção, os tripomastigotas invadem determinados tecidos e se diferenciam em amastigotas, que então se reproduzem no interior das células. Os amastigotas se convertem novamente em tripomastigotas, que rompem as células infectadas e ganham a corrente sanguínea. O ciclo de vida do parasita é ilustrado na figura 3.

Figura 3- Esquemática do ciclo de vida do *T. cruzi*.

Fonte: Adaptado de CLAYTON.¹¹

1.2.2 Aspectos Clínicos da doença de Chagas

A doença de Chagas pode ser dividida em dois estágios: (i) fase aguda, que ocorre logo após a infecção; e, (ii) fase crônica, que fica com o indivíduo pelo resto de sua vida. A fase aguda é normalmente assintomática, entretanto, sintomas como febre, diarreia e vômitos ocorrem em alguns casos. Em aproximadamente 5-10% dos casos sintomáticos, ocorrem processos inflamatórios potencialmente fatais, como miocardite grave ou meningoencefalite.¹³ Grande parte dos indivíduos infectados com o *T. cruzi* não desenvolve os sintomas clínicos da doença e, portanto, são classificados como portadores da forma crônica indeterminada. No entanto, uma parcela dos pacientes desenvolve problemas cardíacos e digestivos sérios, como hipertrofia do miocárdio, megaesôfago e megacólon.

1.3 Descoberta de Fármacos

A descoberta de novos fármacos é um processo bastante longo e complexo, que requer diversas conexões multidisciplinares em P&D, explorando os avanços científicos e tecnológicos para a geração de inovação. A química medicinal possui papel fundamental no estabelecimento de interfaces eficazes para a integração de diferentes áreas, principalmente: química, biologia, física, bioquímica e farmacologia. A química medicinal se dedica a uma variedade de estudos, incluindo o descobrimento, o desenvolvimento e a interpretação do mecanismo de ação de compostos bioativos, bem como a avaliação da sua toxicidade e de suas propriedades farmacocinéticas (e.g., absorção, distribuição, metabolismo e excreção - ADME).¹⁴ A natureza multidisciplinar destes estudos é uma característica que atribui aos métodos de química medicinal grande versatilidade em P&D de fármacos, tanto na indústria, quanto na academia.

O desenvolvimento de um novo fármaco consiste em duas etapas: (i) pré-clínica e (ii) clínica. Durante a fase pré-clínica são realizados estudos relacionados as bases fisiopatológicas da doença e a validação de alvos moleculares. Nas etapas iniciais de P&D também são identificadas moléculas capazes de modular a atividade do alvo molecular e produzir o efeito biológico desejado. Estas moléculas passam por várias etapas de otimização de propriedades farmacodinâmicas e farmacocinéticas até a descoberta de novas entidades químicas (NCE, na sigla inglesa para *new chemical entities*). Nas fases clínicas de desenvolvimento, as NCEs são submetidas a testes em humanos para avaliar sua eficácia e toxicidade.¹⁵

1.4 Modelagem Molecular e Planejamento de Fármacos

A utilização de métodos computacionais no processo de P&D de novos fármacos tem ganho amplo destaque nos últimos anos. A grande disponibilidade de técnicas úteis, aliada a enorme capacidade de processamento de dados e a algoritmos cada vez mais sofisticados, tem impulsionado o crescimento do interesse

dos pesquisadores no uso destas estratégias em projetos de química medicinal e planejamento de fármacos.¹⁶⁻¹⁸

O conhecimento da estrutura tridimensional de macromoléculas biológicas é capital para a aplicação de métodos de planejamento de fármacos baseadas na estrutura do receptor (SBDD, na sigla inglesa para *structure-based drug design*).¹⁹ Dentre as técnicas de SBDD, se destacam a triagem virtual (VS, na sigla inglesa para virtual screening) e a docagem molecular.¹⁴ Em contrapartida, os métodos de síntese orgânica, química combinatória, e de avaliação biológica em alta escala, levaram a aumento notável do número de novas moléculas com atividade biológica determinada. As informações sobre as moléculas bioativas são fundamentais para a aplicação de métodos de planejamento de fármacos baseados na estrutura do ligante (LBDD, na sigla inglesa para *ligand-based drug design*).¹⁸ No LBDD, características moleculares de séries de compostos são identificadas e correlacionadas com sua atividade biológica.¹⁴ Os principais métodos compreendem a geração de modelos farmacofóricos e o desenvolvimento de relações quantitativas entre a estrutura e atividade (QSAR, na sigla inglesa para *quantitative-structure-activity relationships*).¹⁹ Estas diferentes abordagens de modelagem molecular são utilizadas para simular e prever o comportamento de sistemas químico-biológicos e sempre que associadas eficazmente ao trabalho experimental, contribuem de modo significativo em P&D de novos fármacos.

1.5 Relações Quantitativas entre Estrutura e Atividade

A modelagem de QSAR é uma estratégia amplamente utilizada em química medicinal. Os métodos de QSAR baseiam-se na premissa de que os valores de uma determinada propriedade biológica (propriedade alvo) são determinados por características estruturais específicas (descritores moleculares). Consequentemente, variações na propriedade alvo de uma série de moléculas estão associadas a diferenças entre os seus descritores. Valores de atividade biológica (e.g., potência, afinidade) e propriedades farmacocinéticas estão entre as propriedades alvo mais utilizadas em QSAR.²⁰ As propriedades biológicas mais comumente usadas são

obtidas por meio de ensaios experimentais padrões.^{21,22} Os descritores moleculares, por sua vez, podem ser características como o coeficiente de participação (P); a área de superfície polar; as impressões digitais moleculares; e os campos moleculares de interação.¹⁹ Os descritores moleculares podem ser obtidos experimentalmente ou por meio de cálculos computacionais. Todas as metodologias de QSAR fazem uso de modelos matemáticos para quantificar a variação dos valores da propriedade alvo em função dos descritores moleculares.

Uma grande vantagem dos modelos de QSAR no planejamento de fármacos é a sua capacidade de prever a atividade de compostos ainda não testados ou sintetizados, auxiliando o processo de otimização molecular.²³

1.5.1 Validação dos Modelos de QSAR

A validação dos modelos de QSAR é uma etapa essencial no planejamento de novas moléculas. Diferentes parâmetros estatísticos são gerados e analisados como critérios de avaliação da qualidade dos modelos.^{24,25} Os valores da propriedade alvo obtidos experimentalmente são comparados com os valores calculados pelo modelo. Quanto maior a proximidade entre estes valores, mais consistente é o modelo. Os principais parâmetros estatísticos avaliados são o coeficiente de correlação (r^2); o coeficiente de correlação com validação cruzada (q^2); e o coeficiente de correlação preditivo (r^2_{pred}). Estes parâmetros avaliam o poder de correlação entre os dados, bem como a capacidade preditiva do modelo para novas moléculas. Valores de r^2 acima de 0,6 e valores de q^2 e r^2_{pred} acima de 0,5 são indicativos de modelos de QSAR com elevada capacidade preditiva. A tabela 1 apresenta informações sobre estes parâmetros estatísticos.

Tabela 1- Principais parâmetros estatísticos utilizados para se avaliar a qualidade de um modelo de QSAR.

Parâmetro	Tipo de validação	Conjunto utilizado	Fórmula
r^2	Interna	Treino	$r^2 = 1 - \frac{\sum_i (y_{ei} - y_{pi})^2}{\sum_i (y_{ei} - \langle y_e \rangle)^2}$
q^2	Cruzada	Treino	$q^2 = 1 - \frac{\sum_i (y_{ei} - y_{pi})^2}{\sum_i (y_{ei} - \langle y_e \rangle)^2}$
r_{pred}^2	Externa	Teste	$r_{pred}^2 = 1 - \frac{\sum_i (y_{ei} - y_{pi})^2}{\sum_i (y_{ei} - \langle y_e \rangle)^2}$

y_{ei} , valores experimentais da propriedade alvo; y_{pi} , valores preditos da propriedade alvo; $\langle y_e \rangle$, média dos valores experimentais do conjunto treino.

Fonte: Elaborado pelo autor.

Durante o processo de geração dos modelos, as moléculas investigadas (conjunto de dados) são divididas em dois conjuntos: treinamento e teste. O conjunto treinamento é utilizado na construção do modelo e na determinação dos valores de q^2 e r^2 . O conjunto teste, por sua vez, não é utilizado na construção inicial do modelo, sendo empregado posteriormente para avaliar a sua habilidade preditiva (frente a novas moléculas). A capacidade de predição externa é dada pelo valor de r_{pred}^2 .

O parâmetro de validação cruzada (q^2) pode ser avaliado de duas maneiras: (i) pelo método “deixe um de fora” (LOO, na sigla inglesa para *leave-one-out*), em que cada molécula é retirada pelo menos uma vez do conjunto treinamento e sua atividade é determinada pelo modelo gerado; e, (ii) pelo método “deixe muitos de fora” (LMO, na sigla inglesa para *leave-many-out*), em que o conjunto treinamento é dividido em subconjuntos, que seletivamente são excluídos, um a um, para determinação dos valores da propriedade alvo pelo modelo gerado.

Embora os parâmetros estatísticos afirmem a capacidade de correlação e de predição dos modelos, eles não avaliam a redundância dos dados utilizados na construção de modelos de QSAR.^{26,27} Informações redundantes podem levar a um aumento artificial nos valores de q^2 e r^2 . Ainda que dados redundantes sejam capazes de gerar bons coeficientes de correlação, os modelos resultantes possuem precário significado físico.²⁶ Técnicas de randomização da propriedade alvo são

utilizadas para avaliar a possibilidade de redundância.²⁷ Um dos métodos mais conhecidos aplica uma mistura progressiva (do inglês *progressive scrambling*) dos valores da propriedade biológica.²⁸ Após cada ciclo de mistura, os valores de q^2 e r^2 são calculados e a dependência de q^2 em relação a r^2 é dada pelo parâmetro dq^2/dr^2 . Valores de dq^2/dr^2 próximos a um são esperados para conjuntos com pouca redundância. Outro parâmetro fornecido é o Q^2 , que estima a capacidade preditiva do modelo após a eliminação dos dados redundantes.

1.5.2 QSAR Bidimensional - HQSAR

O método holograma QSAR (HQSAR, na sigla inglesa para *Hologram Quantitative Structure-Activity Relationships*) se baseia no conceito de impressão digital molecular. As impressões digitais codificam a estrutura molecular bidimensional em notações numéricas.²⁹ No método HQSAR a estrutura molecular é codificada em uma sequência numérica denominada holograma molecular. Os hologramas moleculares são utilizados como descritores e correlacionados com a propriedade alvo para gerar os modelos de QSAR.³⁰

No processo de construção do holograma molecular, as moléculas são divididas em fragmentos, que podem ser lineares, cíclicos, ramificados e se sobreporem uns aos outros (figura 4). Os fragmentos são gerados seguindo critérios de tamanho mínimo e máximo de átomos, e são distinguidos um dos outros de acordo com diversos parâmetros: tipo de átomo; tipo de ligação; conectividade; átomos de hidrogênio; quiralidade e grupos doadores e aceptores de ligação de hidrogênio (tabela 2). De acordo com estes critérios todos os fragmentos moleculares são gerados e distribuídos ao longo do holograma molecular. Fragmentos semelhantes que ocorrem mais de uma vez são sempre alocados na mesma posição do holograma. Desta forma, os valores numéricos atribuídos a cada posição do holograma representam a quantidade de vezes que determinado fragmento ocorre na molécula.

Figura 4- Etapas da construção de um modelo de HQSAR.

Fonte: Elaborado pelo autor.

Tabela 2- Parâmetros de distinção de fragmentos no método HQSAR.

Parâmetro de distinção de fragmento	Descrição
Átomos (A)	Distingue os fragmentos com base nos tipos de átomos.
Ligações (B)	Os fragmentos são diferenciados com base nos tipos de ligação: simples; dupla; tripla ou aromática.
Conectividade (C)	Os fragmentos são diferenciados com base na hibridização dos átomos presentes no fragmento.
Hidrogênios (H)	Considera os hidrogênios na criação dos fragmentos e na distinção entre eles.
Quiralidade (Ch)	Os fragmentos são diferenciados com base na quiralidade.
Doador e Aceptor (DA)	Distingue os fragmentos com base na presença de grupos aceptores ou doadores de ligação de hidrogênio.

Fonte: Elaborado pelo autor.

Após a construção dos hologramas moleculares, o método estatístico de regressão por mínimos quadrados parciais (PLS, na sigla inglesa para *Partial Least Squares*) correlaciona os valores da propriedade biológica com as informações contidas nos hologramas moleculares.³¹ O modelo gerado é utilizado na predição da

propriedade biológica de novas moléculas. Além disso, os modelos de HQSAR são capazes de identificar características moleculares que determinam a variação da propriedade biológica. Para esta finalidade são gerados mapas de contribuição 2D, que empregam um código de cores para identificar os fragmentos moleculares que influenciam de forma positiva ou negativa a propriedade alvo.

1.5.3 QSAR Tridimensional – CoMFA e CoMSIA

Os métodos de QSAR tridimensionais (QSAR 3D) são bastante difundidos no planejamento de fármacos.²⁰ Estes métodos assumem que as variações na propriedade biológica resultam de interações intermoleculares de natureza 3D com um receptor alvo. Entre as principais metodologias de QSAR 3D, destacam-se a análise comparativa de campos moleculares (CoMFA, na sigla inglesa para *comparative molecular field analysis*) e a análise comparativa de índices de similaridade molecular (CoMSIA, na sigla inglesa para *comparative molecular similarity indices analysis*).^{32,33} Nesses dois métodos, as moléculas são alinhadas e inseridas em uma grade tridimensional reticulada. No método CoMFA, a energia de interação entre uma sonda virtual e os átomos das moléculas é calculada em cada ponto de interseção da grade. Os valores da energia de interação eletrostática e estereoquímica em cada ponto são utilizados como descritores moleculares. O método CoMFA utiliza os potenciais de Lennard-Jones e Coulomb para calcular os campos estereoquímicos (van der Waals) e eletrostáticos, respectivamente.

No método CoMSIA, a sonda virtual é utilizada para calcular índices de similaridade em cada ponto da grade, levando em consideração a distância entre a sonda e os átomos. Os índices de similaridade calculados para cada ponto são utilizados como descritores moleculares. Uma função gaussiana é utilizada para calcular estes índices de similaridade. Além dos potenciais eletrostáticos e estereoquímicos, o método CoMSIA inclui campos de interação hidrofóbicos e considera a presença de interações de hidrogênio. A figura 5 ilustra a construção de modelos CoMFA e CoMSIA.

Figura 5- Esquema de construção de modelos CoMFA e CoMSIA. As moléculas do conjunto de dados são alocadas em uma grade tridimensional e uma sonda é utilizada como átomo de prova.

Moléculas	Atividade	Energia de Interação CoMFA		Índices de Similaridade CoMSIA (A)	
		van der Waals	Eletrostática	(A) ponto 1	(A) ponto 2
1	7,77	6,80	8,99	8,88	6,45
2	3,44	5,44	12,56	5,23	5,80
3	1,98	2,57	6,48	1,34	5,15

↓ PLS
Modelo CoMFA

↓ PLS
Modelo CoMSIA

Fonte: Elaborado pelo autor.

Os métodos CoMFA e CoMSIA produzem mapas de contorno que indicam regiões do espaço em que os descritores moleculares são fortemente correlacionados com a variação da propriedade biológica. Desta forma, essas regiões podem ser exploradas adequadamente no planejamento de novos compostos com propriedades moleculares superiores.

1.5.4 Alinhamento Estrutural

Os métodos de QSAR 3D são sensíveis ao alinhamento molecular do conjunto de dados.³² Diversas estratégias podem ser aplicadas no processo de alinhamento estrutural. Destacam-se: (i) a sobreposição de estruturas de mínima energia; (ii) a utilização da máxima subestrutura comum entre as moléculas; (iii) o alinhamento com o auxílio de farmacóforos; e, (iv) a docagem molecular.³⁴⁻³⁶ O alinhamento com base na mínima energia emprega a sobreposição das estruturas minimizadas de cada molécula de maneira rígida, utilizando como referência um fragmento molecular comum entre as moléculas. Apesar desta metodologia gerar bons alinhamentos, a ausência de liberdade rotacional dos diedros dificulta sua aplicação em estudos de QSAR. Por outro lado, o alinhamento baseado na máxima subestrutura comum emprega uma molécula como referência. Os demais compostos são sobrepostos sobre essa referência utilizando a máxima subestrutura comum como critério de alinhamento. Por sua vez, as técnicas de alinhamento com farmacóforos identificam os grupos funcionais responsáveis pela propriedade biológica de uma série de compostos. Elas se baseiam no princípio de que se um conjunto de compostos que possui a mesma atividade biológica, deve interagir de forma semelhante com o alvo molecular, pelo mesmo mecanismo de ação.³⁶⁻³⁹ Desta forma, características moleculares essenciais são exploradas para caracterizar as interações fármaco-receptor predominantes, tais como a presença de grupos hidrofóbicos, aromáticos, ou ainda, doadores e aceptores de interações de hidrogênio. Diversos programas podem realizar esse tipo de alinhamento.³⁹ Podem ser mencionados o algoritmo genético com atribuição linear para alinhamento de um conjunto de hipermoléculas (GALAHAD, na sigla inglesa para *genetic algorithm with linear assignment for hypermolecular alignment of datasets*) e o programa de similaridade com algoritmo genético (GASP, na sigla inglesa para *genetic algorithm similarity program*).^{40,41} Por fim, a docagem molecular é utilizada quando a estrutura tridimensional do alvo biológico é conhecida. Neste método as moléculas são acopladas à cavidade de interação do alvo molecular e alinhadas de acordo com a sua conformação bioativa.

Objetivos

2 Objetivos

O objetivo geral desta dissertação de mestrado é o desenvolvimento de modelos de QSAR para uma série de derivados piridínicos com potente ação *in vitro* anti-*T. cruzi*. Os objetivos específicos incluem:

- ✓ Padronização de um conjunto de dados de compostos ativos contra o *Trypanosoma cruzi* para a geração de modelos de QSAR.
- ✓ Desenvolvimento de modelos de QSAR utilizando os métodos de HQSAR, CoMFA e CoMSIA.
- ✓ Avaliação da robustez estatística dos modelos gerados, através de sua consistência de correlação interna e de sua capacidade preditiva externa.
- ✓ Identificação de padrões estruturais relevantes para a atividade biológica por meio dos mapas 2D de contribuição (HQSAR) e 3D de contorno (CoMFA e CoMSIA).

Materiais e Métodos

3 Materiais e Métodos

Todos os trabalhos de modelagem de QSAR foram realizados no Laboratório de Química Medicinal e Computacional (LQMC) do Instituto de Física de São Carlos (IFSC) da Universidade de São Paulo (USP). Computadores Dell XPS, que utilizam o sistema operacional Linux CentOS, foram empregados nos trabalhos de modelagem molecular. Os modelos de QSAR foram gerados e analisados com o programa computacional SYBYL-X 2.0 (Tripos, St. Louis, EUA).

3.1 O Conjunto de Dados

Nesta dissertação de mestrado, um conjunto de derivados piridínicos foi utilizado para a geração dos modelos de QSAR. A estrutura destas moléculas é inspirada em um composto com potente atividade tripanocida, identificado pela TDR/WHO em triagens fenotípicas em larga escala. Este composto foi utilizado como referência na síntese e avaliação biológica de uma série de análogos em parceria com o laboratório coordenado pelo Prof. Dr. Luiz Carlos Dias do Instituto de Química da UNICAMP. A série de compostos foi avaliada no LQMC-IFSC-USP por meio de ensaios *in vitro* (estes dados serão publicados em breve). Os ensaios biológicos foram realizados em condições padronizadas utilizando a cepa de *T. cruzi* Tulahuen *LacZ*, que expressa de forma constitutiva a enzima β -galactosidase.^{42,43}

O conjunto de dados consiste em 130 compostos cuja atividade anti-*T. cruzi* é expressa como a concentração necessária para inibir 50% a reprodução do parasita em cultura (IC_{50}). Os valores de IC_{50} variam de 0,30 a 100 μ M, um fator de potência de 333,3 vezes. Os valores de potência biológica foram convertidos para escala logarítmica (pIC_{50}) para adequá-los ao processo de modelagem de QSAR.

Do conjunto total contendo 130 moléculas, 80% (104 compostos) foram utilizados como série treinamento para a construção dos modelos. O restante do conjunto (26 moléculas) foi empregado exclusivamente como série teste. A distribuição das moléculas em relação aos valores de pIC_{50} foi utilizada como critério

para a seleção das séries treinamento e teste. A escolha das moléculas foi efetuada de forma a atribuir aos dois conjuntos uma ampla faixa de distribuição ao longo dos valores de atividade.

Esta série de moléculas possui características adequadas para a realização de trabalhos de QSAR e de otimização molecular. A propriedade biológica foi obtida por meio de ensaios padronizados e validados em nosso laboratório. A diversidade estrutural, caracterizada pelas estruturas químicas de todos os compostos do conjunto de dados, pode ser considerada bastante satisfatória, bem como a distribuição dos valores da potência biológica. Portanto, esse conjunto de dados foi utilizado para a construção de modelos de HQSAR, CoMFA e CoMSIA.

3.2 Estruturas Tridimensionais

Os parâmetros moleculares padrões do programa SYBYL-X 2.0 foram aplicados à construção das estruturas tridimensionais das moléculas do conjunto de dados. A minimização energética destas estruturas foi configurada de acordo com a seguinte parametrização: método Powell (gradiente conjugado); mudança mínima de energia de 0,005 kcal/molÅ; cargas Gasteiger-Huckel; constante dielétrica igual a 80; campo de força Tripos; e número máximo de passos de minimização igual a 1000.

3.3 Alinhamento Molecular

A realização de estudos de QSAR 3D exige que todas as moléculas do conjunto de dados sejam alinhadas no espaço cartesiano. Este procedimento é necessário para que os descritores CoMFA (energia de interação) e CoMSIA (índices de similaridade), calculados em cada ponto da grade, possam ser comparados entre as diferentes moléculas do conjunto de dados. À medida que os resultados destes métodos 3D são sensíveis ao alinhamento, três estratégias diferentes foram utilizadas: (i) alinhamento baseado nas estruturas de mínima

energia; (ii) alinhamento baseado na máxima subestrutura comum; e (iii) alinhamento por superposição de grupos farmacofóricos.

No alinhamento baseado nas estruturas de mínima energia, as moléculas do conjunto de dados foram alinhadas de forma rígida sobre uma subestrutura comum a todos os compostos. Os ângulos diedros foram mantidos com seus valores originais, permitindo-se apenas a rotação e a translação das moléculas.

O alinhamento baseado na máxima subestrutura comum foi executado através do programa Distill Rigid. As moléculas do conjunto de dados foram alinhadas de forma flexível sobre uma subestrutura comum a todos os compostos. Os valores dos ângulos diedros da estrutura de mínima energia do composto mais potente da série (molécula **85**) foram utilizados como referência para o alinhamento molecular.

O alinhamento por superposição de grupos farmacofóricos foi realizado por meio do programa GALAHAD. O método extrai as características farmacofóricas das moléculas e posteriormente as sobrepõe de acordo com o arranjo tridimensional das características identificadas. Uma população de 500 indivíduos foi utilizada para a aplicação dos operadores genéticos, permitindo-se um número máximo de 50 gerações. Todas as propriedades farmacofóricas implementadas no GALAHAD foram utilizadas (grupos hidrofóbicos, doadores e aceptores de interações de hidrogênio, anéis de seis átomos, centros negativos e átomos de nitrogênio positivos). O modo de alinhamento mútuo foi aplicado, ou seja, nenhuma molécula foi utilizada como referência.

3.4 HQSAR

Modelos de HQSAR foram gerados para o conjunto de dados de acordo com o algoritmo padrão implementado no programa SYBYL-X 2.0. Os hologramas moleculares foram originados com a utilização dos comprimentos padrões: 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, e 401. Diferentes combinações de parâmetros de distinção de fragmentos foram utilizadas pelo algoritmo para gerar os

diversos fragmentos. Os modelos também foram testados quanto a sua sensibilidade à variação do tamanho dos fragmentos. Para todos os modelos gerados foi empregado um número máximo de oito componentes. O algoritmo HQSAR utilizou os valores de r^2 como critério para reportar a melhor combinação de parâmetros.⁴⁴

3.5 CoMFA

Modelos CoMFA foram gerados para o conjunto de dados de acordo com o algoritmo padrão implementado no SYBYL-X 2.0.⁴⁴ Uma grade tridimensional com espaços de 2 Å entre cada retículo foi utilizada, estendendo-se por 4 Å além das extremidades das moléculas. Um átomo de carbono sp^3 com carga +1 foi utilizado como átomo de prova. A energia de interação entre as moléculas e o átomo de prova foi calculada através dos potenciais de Lennard-Jones e Coulomb, aplicando-se a parametrização padrão do campo de força Tripos. Um valor de 30 kcal/mol foi utilizado como valor de corte para a energia de interação eletrostática e estereoquímica. Uma forma diferenciada do potencial de Coulomb em que a energia de interação diminui proporcionalmente ao quadrado da distância r entre a sonda e os átomos foi utilizada. O método de focalização de região foi aplicado para aumentar o peso dos descritores que contribuem de forma mais consistente para a variação da propriedade biológica. A técnica de focalização utilizada foi o ajuste dos coeficientes de desvio padrão (SDC, na sigla inglesa para *standard deviation coefficient*).

Os mapas de contorno CoMFA foram gerados calculando-se o produto do desvio padrão das energias de interação pelos coeficientes PLS. Os pontos da grade em que este produto resultou em valores acima de 80% ou abaixo de 20% do valor máximo calculado, foram utilizados para a construção dos mapas.

3.6 CoMSIA

Modelos CoMSIA foram gerados para o conjunto de dados de acordo com o algoritmo padrão implementado no SYBYL-X 2.0. Uma grade tridimensional foi utilizada, com as mesmas configurações da empregada na modelagem CoMFA. A sonda virtual também foi a mesma utilizada na construção dos modelos CoMFA. Os índices de similaridade em cada ponto da grade foram calculados através de uma função gaussiana. Diferentes valores do fator de atenuação da função foram aplicados. O recurso de filtragem de coluna também foi aplicado aos modelos CoMSIA para eliminar descritores com valores de energia abaixo de 2 kcal/mol. O método SDC de focalização de região foi utilizado para refinar os modelos.

Todos os campos moleculares implementados no método CoMSIA foram empregados – campos hidrofóbicos, estereoquímicos, eletrostáticos, e referentes a doadores e aceptores de interações de hidrogênio. Os mapas de contorno foram gerados calculando-se o produto do desvio padrão dos índices de similaridade pelos coeficientes PLS. Os pontos da grade em que o resultado deste produto assumiu valores acima de 80% ou abaixo de 20% do valor máximo foram utilizados para a construção dos mapas CoMSIA.⁴⁶

3.7 Validação Interna e Externa dos Modelos

A consistência estatística e a capacidade de predição externa foram avaliadas para os melhores modelos.⁴⁷ O conjunto teste foi completamente excluído da etapa de geração dos modelos. O método PLS com validação cruzada LOO foi utilizado na determinação do número ótimo de componentes e dos valores de q^2 . A divisão do conjunto treinamento em vários grupos (validação LMO) foi empregada para avaliar a estabilidade dos modelos com a variação de dados. A possibilidade de redundância dos modelos gerados foi aferida por meio da permutação dos valores da propriedade biológica dos compostos. Após a realização de todos esses procedimentos de validação interna, os modelos foram avaliados quanto à sua

capacidade de predição (validação externa), empregando-se um conjunto teste com 26 compostos, selecionados de acordo com a metodologia descrita na seção 3.1.

Resultados e Discussão

4 Resultados e Discussão

4.1 O Conjunto de Dados

Um conjunto de dados constituído de 130 derivados piridínicos foi utilizado na construção dos modelos de HQSAR, CoMFA e CoMSIA. A figura 6 mostra o fragmento farmacofórico comum a esta série, o qual é essencial para a conservação da atividade biológica. Os diferentes substituintes explorados nas posições destacadas (R^1 , R^2 e R^3) atribuem considerável diversidade química ao conjunto. A tabela 3 ilustra a estrutura dos compostos e seus respectivos valores de pIC_{50} .

Na posição R^1 foram explorados grupos volumosos como substituintes do anel piridínico. Dentre os principais substituintes podemos destacar: metila, metoxifenila, *tert*-butila, fenila, furano, nitrofenila e fluorofenila. Em R^2 destacam-se os substituintes metila, fenila, metoxifenila, naftaleno e metila. Maior diversidade estrutural foi conferida à posição R^3 , como pode ser observado na tabela 3. Grande parte dos compostos possui um espaçador contendo um ou dois átomos de carbono entre o enxofre e os substituintes em R^3 .

Figura 6 – Farmacóforo característico da série de compostos utilizada na modelagem de QSAR.

Fonte: Elaborado pelo autor.

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
1		5,06	2*		4,30
3		4,30	4		4,30
5		4,30	6		4,30
7		4,30	8*		4,46
9		4,47	10		4,74
11		4,41	12		4,49
13*		4,49	14		4,49

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
15		5,15	16		4,95
17		4,47	18		4,84
19*		4,76	20		4,30
21*		4,65	22*		4,86
23		4,32	24		4,30
25		4,82	26		4,76
27		4,48	28		4,30

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
29		4,30	30*		4,30
31		4,30	32		4,30
33		4,30	34*		4,94
35		4,30			4,39
37		5,03	38*		4,30
39		4,30	40		4,00
41		4,00	42		4,00

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
43*		4,00	44		4,00
45		4,00	46		4,00
47		4,00	48		4,55
49		4,30	50		5,00
51		4,92	52*		5,22
53		4,76	54		4,00
55		4,00	56		4,00

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
57		4,00	58		4,74
59		5,09	60		4,00
61		4,00	62		4,00
63		4,00	64*		4,00
65		4,00	66		4,00
67		4,00	68		6,00
69		4,00	70		4,00
71*		4,00	72		4,09

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
73		5,18	74*		4,00
75		4,00	76*		4,76
77		4,32	78		4,00
79		4,00	80		4,00
81		4,00	82		4,00
83		4,00	84		4,00
85		6,52	86*		6,48
87		6,45	88		6,05
89		5,88	90		5,46
91		5,30	92		4,45

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
93*		5,76	94		4,00
95*		5,85	96		4,00
97*		4,85	98		4,00
99		5,09	100		4,50
101		5,79	102*		5,60
103		5,30	104		5,88
105		5,31	106		5,39
107		5,09	108		5,49
109*		5,30	110		5,00
111		5,69	112		4,60

(Continua)

(Continuação)

Tabela 3 – Estrutura das moléculas do conjunto de dados e seus valores de pIC_{50} .

Molécula	Estrutura	pIC_{50}	Molécula	Estrutura	pIC_{50}
113		4,58	114		5,45
115		4,48	116*		4,00
117		4,15	118		4,00
119		4,00	120		4,00
121*		4,74	122*		4,00
123		4,00	124		4,00
125		4,00	126*		4,74
127*		5,63	128		6,00
129		4,00	130		4,00

*Moléculas da série teste.

Fonte: Elaborado pelo autor.

Os valores de pIC_{50} variam de 4,00 (compostos menos potentes) a 6,52 (molécula mais potente). A figura 7 mostra a distribuição dos valores de pIC_{50} . Com base nestes dados, as moléculas foram divididas em dois conjuntos, denominados treinamento e teste. A seleção foi realizada de forma que os dois conjuntos apresentassem uma boa distribuição ao longo do intervalo de atividade. Este equilíbrio entre os conjuntos treinamento e teste é um aspecto crítico na produção de modelos consistentes. Os seguintes compostos foram selecionados para integrar o conjunto teste: **2, 8, 13, 19, 21, 22, 30, 34, 38, 43, 52, 64, 71, 74, 76, 86, 93, 95, 97, 102, 109, 116, 121, 122, 126 e 127**. O restante das moléculas foi utilizado como conjunto treinamento.

Figura 7 - Distribuição dos valores da propriedade biológica nos conjuntos teste, treinamento e no conjunto total.

Fonte: Elaborado pelo autor.

4.2 HQSAR

Os modelos de HQSAR foram construídos utilizando-se os seguintes tamanhos de holograma: 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401. Diversas combinações dos parâmetros de distinção de fragmentos foram utilizadas para estabelecer os modelos iniciais: A/B, A/B/C, A/B/C/H, A/C, A/DA, A/B/H, A/B/DA, A/B/H/DA, A/B/C/DA, A/B/C/H/DA, A/C/H, A/C/DA, B/DA, B/C/H/DA. Diferentes tamanhos dos fragmentos (de 2 a 5 átomos até 7 a 10 átomos) foram empregados para todas estas combinações. Os modelos foram avaliados inicialmente pelo método LOO. Os resultados obtidos apenas para o tamanho dos fragmentos de 7 a 10 átomos são exibidos na tabela 4.

Tabela 4 - Resultados de HQSAR para diferentes parâmetros de distinção de fragmentos, utilizando o tamanho dos fragmentos de 7 a 10 átomos.

Distinção de fragmentos	q^2	SEP	r^2	SEE	HL	N
A/B	0,70	0,35	0,82	0,27	199	6
A/B/C	0,67	0,37	0,84	0,26	257	6
A/B/C/H	0,62	0,40	0,82	0,27	401	6
A/C	0,68	0,36	0,83	0,26	257	6
A/DA	0,65	0,38	0,80	0,28	353	6
A/B/H	0,61	0,40	0,79	0,29	151	6
A/B/DA	0,62	0,39	0,82	0,27	401	6
A/B/H/DA	0,55	0,43	0,79	0,29	257	6
A/B/C/DA	0,64	0,39	0,84	0,25	307	6
A/B/C/H/DA	0,57	0,42	0,79	0,29	353	6
A/C/H	0,64	0,39	0,83	0,26	401	6
A/C/DA	0,67	0,37	0,83	0,26	353	6
B/DA	0,66	0,38	0,81	0,27	199	6
B/C/H/DA	0,62	0,40	0,83	0,26	199	6

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; HL, Comprimento do holograma; N, Número de componentes. A, átomos; B, ligações; C, conectividade; H, hidrogênios; DA, doador/aceptor de ligação de hidrogênio.

Fonte: Elaborado pelo autor.

Os melhores modelos obtidos nesta etapa inicial foram selecionados com base nos valores de q^2 . A seguir, estes modelos foram avaliados quanto a sua

capacidade de predição para o conjunto teste, por meio do indicador estatístico r^2_{pred} . A tabela 5 mostra os resultados obtidos para este conjunto de modelos.

Tabela 5 - Modelos estatísticos gerados pelo método de HQSAR.

Distinção de fragmentos	Tamanho de fragmento	q^2	SEP	r^2	SEE	HL	N	r^2_{pred}
A/B	5-8	0,67	0,37	0,82	0,27	401	6	0,67
A/B/C	5-8	0,66	0,37	0,81	0,28	61	6	0,62
B/C/H/DA	5-8	0,65	0,38	0,84	0,25	199	6	0,73
A/B	6-9	0,68	0,36	0,82	0,27	257	6	0,65
A/B/C	6-9	0,66	0,37	0,74	0,32	257	3	0,73
A/C	6-9	0,66	0,38	0,83	0,26	353	6	0,77
A/B/C/DA	6-9	0,65	0,38	0,83	0,27	199	6	0,78
A/C/DA	6-9	0,66	0,38	0,81	0,28	353	6	0,75
A/B	7-10	0,70	0,35	0,82	0,27	199	6	0,78
A/B/C	7-10	0,67	0,37	0,84	0,26	257	6	0,75
A/C	7-10	0,68	0,36	0,83	0,26	257	6	0,73
A/DA	7-10	0,65	0,38	0,80	0,28	353	6	0,69
A/C/DA	7-10	0,67	0,37	0,83	0,26	353	6	0,79
B/DA	7-10	0,66	0,38	0,81	0,27	199	6	0,66

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; HL, Comprimento do holograma; N, Número de componentes; r^2_{pred} , coeficiente de correlação preditivo; A, átomos; B, ligações; C, conectividade; H, hidrogênios; DA, doador/aceptor de ligação de hidrogênio.

Fonte: Elaborado pelo autor.

Como pode ser observado, o modelo A/C/DA com tamanho dos fragmento de 7 a 10 átomos apresentou a maior capacidade preditiva ($r^2_{\text{pred}} = 0,79$). A tabela 6 mostra os resultados para os diferentes tamanhos dos fragmentos, utilizando a distinção de fragmentos A/C/DA. Pode-se observar que os valores de q^2 aumentaram continuamente até o tamanho dos fragmentos ótimo (7-10). A partir deste ponto, os valores começaram a decrescer, ilustrando a importância de se explorar diferentes tamanhos dos fragmentos.

Tabela 6 - Resultados de HQSAR para diferentes tamanhos dos fragmentos (modelo A/C/DA).

Tamanho dos fragmentos	q^2	SEP	r^2	SEE	HL	N
2-5	0,51	0,45	0,68	0,36	59	6
3-6	0,53	0,44	0,72	0,33	257	5
4-7	0,55	0,43	0,77	0,31	353	6
5-8	0,61	0,40	0,77	0,30	97	6
6-9	0,66	0,38	0,81	0,28	353	6
7-10	0,67	0,37	0,83	0,26	353	6
8-11	0,64	0,38	0,83	0,27	257	6
9-12	0,61	0,40	0,82	0,27	401	6

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

Fonte: Elaborado pelo autor.

Para avaliar a estabilidade do melhor modelo (A/C/DA, tabela 5) frente a variação dos dados, o método LMO foi aplicado. Para isso, o conjunto treinamento foi dividido aleatoriamente em cinco e dez grupos de validação. Os resultados são exibidos na tabela 7.

Tabela 7 - Resultados LMO para o modelo A/C/DA com tamanho dos fragmentos de 7 a 10 átomos.

Grupos de validação	q^2	SEP	r^2	SEE	HL	N
5	0,64	0,38	0,81	0,28	257	6
10	0,66	0,37	0,83	0,26	353	6

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

Fonte: Elaborado pelo autor.

Como pode ser observado, os resultados são comparáveis aos obtidos pelo método LOO. Além disso, a possibilidade de redundância do melhor modelo (A/C/DA, tamanho dos fragmentos de 7-10) foi avaliada usando o método de permutação da atividade biológica. O resultado indicou uma baixa redundância do modelo ($dq^2/dr^2 = 1,03$).

Embora os procedimentos de validação interna sejam de fundamental importância na geração de bons modelos de QSAR, os mesmos devem ser avaliados quanto à sua capacidade preditiva para moléculas que não foram utilizadas na construção do modelo (série teste). Na tabela 8 são listados os valores experimentais e preditos de pIC_{50} para os compostos da série teste. Os dados são ilustrados na figura 8 para todo o conjunto de dados. A boa correlação entre os

valores experimentais e preditos de pIC_{50} para a série teste indica a utilidade do modelo de HQSAR para a predição da atividade anti-*T. cruzi* de novas moléculas, dentro da diversidade estrutural estudada.

Tabela 8 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo de HQSAR).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	4,30	4,68	-0,38
8	4,46	4,56	-0,10
13	4,49	4,40	0,09
19	4,76	4,24	0,52
21	4,65	4,70	-0,05
22	4,86	4,66	0,20
30	4,30	4,03	0,27
34	4,94	4,51	0,43
38	4,30	4,51	-0,21
43	4,00	4,04	-0,04
52	5,22	4,73	0,49
64	4,00	4,30	-0,30
71	4,00	4,47	-0,47
74	4,00	4,61	-0,61
76	4,76	4,62	0,14
86	6,48	5,78	0,70
93	5,76	5,76	0,00
95	5,85	5,77	0,08
97	4,85	5,06	-0,21
102	5,60	5,96	-0,36
109	5,30	5,28	0,02
116	4,00	4,05	-0,05
121	4,74	4,63	0,11
122	4,00	4,17	-0,17
126	4,74	4,26	0,48
127	5,63	5,95	-0,32

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

Fonte: Elaborado pelo autor.

Figura 8 - Valores experimentais e valores preditos de pIC_{50} para o melhor modelo de HQSAR.

Fonte: Elaborado pelo autor.

4.2.1 Mapas de contribuição

Os mapas de contribuição de HQSAR indicam os fragmentos moleculares que contribuem positivamente ou negativamente para a atividade biológica. Contribuições positivas são ilustradas em verde e amarelo; em contrapartida, contribuições negativas são ilustradas em vermelho e laranja. A figura 9 mostra os mapas de contribuição para alguns dos compostos mais potentes do conjunto de dados. O anel piridínico é identificado como uma importante contribuição para atividade, assim como o anel benzênico. Os mapas também realçam a contribuição positiva dos átomos de oxigênio nos substituintes R^3 . Estes resultados indicam que

fragmentos contendo aceptores de interações de hidrogênio unidos a sistemas aromáticos contribuem de forma positiva para a atividade biológica.

Figura 9 - Mapas de contribuição de HQSAR para os compostos **88**, **101**, **108** e **89**.

Fonte: Elaborado pelo autor.

No processo de planejamento desta série de compostos, modificações do núcleo piridínico também foram exploradas. Anéis tienopiridínicos foram introduzidos em diversas moléculas (compostos **77**, **78**, **79**, **80**, **94**, **96**, **119** e **126**). Os mapas de contribuição de HQSAR apontam que esta estratégia de modificação, bem como a presença de um grupo amina no núcleo tienopiridina, contribuem negativamente para a atividade biológica (figura 10). Estes indícios são coerentes com a atividade destes derivados já que a introdução destes grupos resultou em perda significativa da atividade determinada experimentalmente.

Figura 10 - Mapas de contribuição de HQSAR para os compostos **78**, **79**, **80** e **96**.

Fonte: Elaborado pelo autor.

4.3 Alinhamento Molecular

Diversas metodologias de alinhamento molecular foram exploradas para a construção dos modelos de QSAR 3D. A proteína alvo desta série de compostos não é conhecida, portanto, foram empregados exclusivamente métodos de LBDD. As seguintes estratégias de alinhamento foram exploradas: (i) máxima subestrutura comum; (ii) superposição de farmacóforos; e, (iii) estruturas de mínima energia.

4.3.1 Alinhamento baseado na máxima subestrutura comum (Alinhamento 1)

Nesta estratégia, os compostos são alinhados sobre uma molécula utilizada como referência. O conjunto de dados é alinhado de forma a maximizar a sobreposição da máxima subestrutura comum a todos os compostos (MCS, na sigla

inglesa para *maximum common substructure*). Os diedros são ajustados de forma a assumir os valores dos ângulos da MCS do composto utilizado como referência. O alinhamento foi realizado com o programa *Distill Rigid*, utilizando como referência o composto mais potente do conjunto (**85**). O alinhamento obtido pelo método MCS (figura 11) ilustra o bom alinhamento dos núcleos piridínicos. Em contrapartida, o alinhamento evidencia a dispersão dos fragmentos que não fazem parte da MCS. Como estas regiões não sofrem rotações dos diedros, este padrão de alinhamento é esperado quando o método MCS é utilizado.

Figura 11 - Alinhamento molecular obtido pelo método MCS.

Fonte: Elaborado pelo autor.

4.3.2 Alinhamento baseado na superposição de farmacóforos (Alinhamento 2)

O programa GALAHAD foi utilizado na obtenção do alinhamento molecular baseado em farmacóforos. Durante o processo de alinhamento molecular foram

observados três grupos de compostos que produziam padrões de alinhamento diferentes entre si. As moléculas de cada grupo apresentaram como característica comum a presença de: um núcleo ciano-piridínico; um anel tienopiridínico; e dois núcleos ciano-piridínicos. Um representante de cada classe foi selecionado (figura 12) e as três moléculas foram alinhadas. Este procedimento com o programa GALAHAD resultou na produção de um modelo farmacofórico e de um alinhamento representativo dos três grupos.

Figura 12 - Compostos utilizados na construção do modelo farmacofórico.

Fonte: Elaborado pelo autor.

O alinhamento realizado resultou em uma boa sobreposição estrutural, como pode ser observado na figura 13. O modelo farmacofórico produzido é caracterizado por dois grupos aromáticos e dois aceptores de interações de hidrogênio. A seguir, os demais compostos do conjunto de dados foram alinhados sobre as conformações obtidas com o GALAHAD. A figura 14 mostra o alinhamento para todo o conjunto de dados.

Figura 13 - Alinhamento gerado com o GALAHAD. As esferas dos anéis, em verde claro, indicam os grupos aromáticos, e as outras duas, em verde, indicam aceptores de interações de hidrogênio.

Fonte: Elaborado pelo autor.

Figura 14 - Alinhamento por farmacóforos.

Fonte: Elaborado pelo autor.

Esta estratégia de alinhamento resultou em uma sobreposição satisfatória de toda a composição molecular. Além do grupo piridínico, os fragmentos que não fazem parte da MCS, ligados às posições R^1 , R^2 e R^3 , foram sobrepostos de

maneira mais precisa em comparação ao alinhamento obtido pelo método anterior (MCS).

4.3.3 Alinhamento baseado nas estruturas de mínima energia (Alinhamento 3)

Nesta estratégia de alinhamento, as estruturas minimizadas, de acordo com o protocolo descrito na seção 3.2, foram alinhadas de forma rígida. Nenhum diedro foi modificado, apenas movimentos de rotação e translação foram permitidos. O núcleo ciano-piridínico foi utilizado como subestrutura comum. A figura 15 mostra o resultado do alinhamento. É possível observar que o anel piridínico foi alinhado satisfatoriamente, porém, os grupos substituintes na posição R^3 ficaram relativamente dispersos. Este resultado é previsível, porque estes substituintes apresentam maior diversidade estrutural e os ângulos diedros são mantidos em suas configurações originais.

Figura 15 - Alinhamento molecular utilizando estruturas de mínima energia.

Fonte: Elaborado pelo autor.

4.4 CoMFA

4.4.1 Modelos CoMFA - Alinhamento 1

Estudos CoMFA foram realizados com o alinhamento baseado na máxima subestrutura comum. O método de validação cruzada LOO foi utilizado inicialmente para avaliar a consistência estatística dos modelos. A ferramenta SDC de focalização de região foi aplicada com o objetivo de aumentar o peso dos descritores

mais significativos para o modelo. A tabela 9 mostra os resultados gerados em função da variação dos valores de SDC.

Tabela 9 - Resultados CoMFA obtidos com o método de focalização de região SDC.

Modelo	q^2	SEP	r^2	SEE	N	SDC
1	0,56	0,44	0,86	0,25	7	-
2	0,67	0,37	0,85	0,25	5	0,3
3	0,69	0,36	0,86	0,25	5	0,5
4	0,70	0,36	0,87	0,24	6	0,7
5	0,69	0,36	0,86	0,25	7	0,9

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

Fonte: Elaborado pelo autor.

O melhor modelo foi obtido ajustando-se o valor de SDC por um fator de 0,7. Posteriormente, os descritores ajustados pelo valor ótimo de SDC foram utilizados para gerar novos modelos. Nesta etapa, foram investigados os efeitos da variação do número de componentes sobre os resultados. A possibilidade de redundância foi avaliada pela técnica de permutação dos valores da atividade biológica. A tabela 10 mostra os resultados obtidos em função do número de componentes. Os dados demonstraram que modelos que possuem de quatro a cinco componentes apresentaram baixa redundância. Como pode ser observado, a utilização de um número de componentes acima de cinco resultou no crescimento dos valores de q^2 e r^2 . No entanto, valores de dq^2/dr^2 maiores que 1,2 indicam modelos excessivamente redundantes. O conjunto destes indicadores apontaram o modelo CoMFA 7 como o melhor para o alinhamento 1.

Tabela 10 - Resultados CoMFA em função da variação do número de componentes (SDC = 0,7; alinhamento 1).

Modelo	q^2	SEP	r^2	SEE	N	dq^2/dr^2
4	0,70	0,36	0,87	0,27	6	1,12
6	0,67	0,36	0,83	0,27	4	1,01
7	0,70	0,36	0,85	0,25	5	1,02
8	0,70	0,36	0,89	0,22	7	1,24
9	0,72	0,35	0,90	0,20	8	1,38

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; N, Número de componentes; dq^2/dr^2 , variação de q^2 em função de r^2 . (Modelo 4, melhor modelo produzido ajustando-se o valor de SDC para 0,7 e o número de componentes para 6, ver tabela 9).

Fonte: Elaborado pelo autor.

O modelo 7 também foi avaliado quanto a sua estabilidade pelo método LMO (obtidos $q^2 = 0,65$ e $0,68$, para cinco e dez grupos de validação, respectivamente). Após a confirmação da sua robustez estatística, o melhor modelo CoMFA foi avaliado quanto a sua capacidade de predição externa. Desta forma, o modelo foi utilizado para prever a atividade do conjunto teste. O poder preditivo do modelo foi confirmado pela obtenção de um valor de $r^2_{\text{pred}} = 0,75$. Os valores experimentais e preditos de pIC_{50} para a série teste são listados na tabela 11. A figura 16 ilustra graficamente a correlação entre os valores experimentais e preditos da atividade biológica para todo o conjunto de dados.

Tabela 11 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 1).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	4,30	4,56	-0,26
8	4,46	4,78	-0,32
13	4,49	4,35	0,14
19	4,76	4,04	0,72
21	4,65	4,78	-0,13
22	4,86	4,82	0,04
30	4,30	3,97	0,33
34	4,94	4,58	0,36
38	4,30	4,59	-0,29
43	4,00	4,16	-0,16
52	5,22	4,79	0,43
64	4,00	4,52	-0,52

(Continua)

(Continuação)

Tabela 11 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 1).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo^a
71	4,00	4,77	-0,77
74	4,00	3,99	0,01
76	4,76	4,75	0,01
86	6,48	6,12	0,36
93	5,76	6,00	-0,24
95	5,85	5,87	-0,02
97	4,85	4,70	0,15
102	5,60	6,20	-0,60
109	5,30	5,05	0,25
116	4,00	3,86	0,14
121	4,74	4,79	-0,05
122	4,00	4,41	-0,41
126	4,74	3,99	0,75
127	5,63	6,24	-0,61

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

Fonte: Elaborado pelo autor.

Figura 16 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 1).

Fonte: Elaborado pelo autor.

4.4.2 Modelos CoMFA - Alinhamento 2

Estudos CoMFA foram realizados com o alinhamento obtido através por superposição de farmacóforos. Os modelos iniciais foram gerados com o mesmo valor de SDC utilizado para o alinhamento 1 ($SDC = 0,7$). A seguir, a influência da variação do número de componentes sobre os resultados foi analisada. A técnica de permutação dos valores da atividade biológica foi aplicada para avaliar a possibilidade de redundância dos modelos gerados. A tabela 12 mostra os indicadores estatísticos obtidos em função da variação do número de componentes.

Tabela 12- Resultados CoMFA obtidos em função da variação do número de componentes (SDC = 0,7; alinhamento 2).

Modelo	q^2	SEP	r^2	SEE	N	dq^2/dr^2
10	0,54	0,44	0,80	0,29	4	0,81
11	0,59	0,42	0,84	0,27	5	0,87
12	0,63	0,40	0,86	0,25	6	0,90
13	0,65	0,39	0,88	0,23	7	1,09
14	0,63	0,40	0,90	0,21	8	1,25

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; N, Número de componentes; dq^2/dr^2 , variação de q^2 em função de r^2 .

Fonte: Elaborado pelo autor.

Os resultados da tabela 12 demonstraram que o modelo gerado com sete componentes apresentou a menor redundância. A estabilidade deste modelo também foi avaliada pelo método LMO (obtidos $q^2 = 0,61$ e $0,64$, para cinco e dez grupos de validação, respectivamente). Esta série de parâmetros estatísticos indicaram o modelo CoMFA 13 como o melhor para o alinhamento 2. A capacidade de predição externa deste modelo foi avaliada através do conjunto teste ($r^2_{\text{pred}} = 0,72$). A tabela 13 mostra os valores experimentais e preditos de pIC_{50} para a série teste. Os resultados para todo o conjunto de dados são ilustrados na figura 17.

Tabela 13 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 2).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	4,30	4,63	-0,33
8	4,46	4,42	0,04
13	4,49	4,52	-0,03
19	4,76	4,00	0,76
21	4,65	4,50	0,15
22	4,86	4,66	0,20
30	4,30	3,68	0,62
34	4,94	4,44	0,50
38	4,30	4,47	-0,17
43	4,00	4,26	-0,26
52	5,22	4,94	0,28
64	4,00	4,13	-0,13
71	4,00	4,52	-0,52
74	4,00	4,62	-0,62
76	4,76	4,91	-0,15

(Continua)

(Continuação)

Tabela 13 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 2).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo^a
86	6,48	5,84	0,64
93	5,76	5,83	-0,07
95	5,85	5,70	0,15
97	4,85	4,73	0,12
102	5,60	5,52	0,08
109	5,30	5,04	0,26
116	4,00	4,41	-0,41
121	4,74	4,27	0,47
122	4,00	4,14	-0,14
126	4,74	4,79	-0,05
127	5,63	5,09	0,54

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

Fonte: Elaborado pelo autor.

Figura 17 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 2).

Fonte: Elaborado pelo autor.

4.4.3 Modelos CoMFA - Alinhamento 3

Modelos CoMFA foram construídos com o alinhamento molecular obtido através do método de estruturas de mínima energia. De maneira análoga aos alinhamentos anteriores, descritores ajustados por um valor de $SDC = 0,7$ foram utilizados na construção dos modelos. Alterações do número de componentes foram aplicadas para gerar os modelos iniciais. Em seguida, a técnica de permutação dos valores da atividade biológica foi aplicada na avaliação da redundância destes modelos. A tabela 14 contém os resultados obtidos de acordo com o número de componentes utilizado.

Tabela 14 - Resultados CoMFA obtidos em função da variação do número de componentes (SDC = 0,7; alinhamento 3).

Modelo	q^2	SEP	r^2	SEE	N	dq^2/dr^2
15	0,48	0,47	0,74	0,33	4	0,74
16	0,54	0,45	0,83	0,27	5	0,87
17	0,59	0,42	0,87	0,24	6	0,90
18	0,61	0,41	0,89	0,22	7	0,93
19	0,62	0,41	0,91	0,20	8	1,01
20	0,61	0,41	0,92	0,18	9	1,04
21	0,60	0,42	0,94	0,17	10	1,30
22	0,61	0,42	0,94	0,16	11	1,39

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; N, Número de componentes; dq^2/dr^2 , variação de q^2 em função de r^2 .

Fonte: Elaborado pelo autor.

Os dados demonstraram que o modelo 19 apresentou os melhores indicadores estatísticos. A estabilidade deste modelo foi avaliada pelo método LMO. Este procedimento resultou em valores de $q^2 = 0,53$ e $0,63$, para cinco e dez grupos de validação, respectivamente. Estes resultados são comparáveis aos obtidos com a técnica LOO. A predição dos valores de atividade para o conjunto teste resultou em um valor de $r^2_{\text{pred}} = 0,64$. A tabela 15 mostra os valores experimentais e preditos de pIC_{50} para a série teste e a figura 18 ilustra os resultados para todo o conjunto de dados.

Tabela 15 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 3).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	4,30	4,26	0,04
8	4,46	4,52	-0,06
13	4,49	4,60	-0,11
19	4,76	3,95	0,81
21	4,65	4,68	-0,03
22	4,86	4,91	-0,05
30	4,30	4,13	0,17
34	4,94	4,49	0,45
38	4,30	4,39	-0,09
43	4,00	4,23	-0,23
52	5,22	4,64	0,58
64	4,00	4,15	-0,15

(Continua)

(Continuação)

Tabela 15 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMFA – alinhamento 3).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo^a
71	4,00	4,55	-0,55
74	4,00	3,88	0,12
76	4,76	4,48	0,28
86	6,48	5,60	0,88
93	5,76	5,36	0,40
95	5,85	6,78	-0,93
97	4,85	4,57	0,28
102	5,60	5,63	-0,03
109	5,30	5,26	0,04
116	4,00	4,39	-0,39
121	4,74	4,41	0,33
122	4,00	4,27	-0,27
126	4,74	4,02	0,72
127	5,63	5,38	0,25

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

Fonte: Elaborado pelo autor.

Figura 18 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMFA (alinhamento 3).

Fonte: Elaborado pelo autor.

Os resultados produzidos demonstraram que diferentes estratégias de alinhamento podem ser utilizadas sem prejuízo significativo para a robustez estatística dos modelos. No entanto, a escolha da metodologia de alinhamento pode afetar significativamente a capacidade preditiva dos modelos gerados. Este aspecto depende fortemente da diversidade estrutural e do grau de flexibilidade das moléculas envolvidas. Para o conjunto de dados investigado neste trabalho, o alinhamento baseado na MCS resultou no modelo CoMFA com maior poder preditivo ($r^2_{\text{pred}} = 0,75$), seguido pelo alinhamento baseado em farmacóforos ($r^2_{\text{pred}} = 0,72$) e pelo alinhamento baseado em estruturas de mínima energia ($r^2_{\text{pred}} = 0,64$). Estes resultados indicam que a introdução de liberdade rotacional para os diedros foi determinante para aumentar a capacidade de predição externa dos modelos CoMFA.

Apesar da diferença entre os valores de r^2_{pred} , todos os modelos apresentaram boa correlação entre os valores experimentais e preditos de pIC_{50} para o conjunto teste. Os resultados obtidos indicam que estes modelos são úteis em química medicinal para a predição da atividade anti-*T. cruzi* de novos compostos dentro da diversidade química explorada.

4.5 CoMSIA

4.5.1 Modelos CoMSIA - Alinhamento 1

Estudos CoMSIA foram realizados com o alinhamento obtido por meio do método MCS. Alterações do fator de atenuação α da função gaussiana, foram utilizadas como estratégia inicial na construção dos modelos CoMSIA. Uma vez identificado um valor ótimo de 0,7 para α , os modelos foram refinados aplicando-se o método SDC de focalização de região e a técnica de filtragem de colunas. A tabela 16 descreve os resultados produzidos em função dos valores de SDC.

Tabela 16 - Resultados CoMSIA obtidos através do método SDC.

Modelo	q^2	SEP	r^2	SEE	N	SDC
23	0,48	0,48	0,88	0,23	7	-
24	0,62	0,40	0,86	0,25	7	0,3
25	0,65	0,39	0,86	0,24	8	0,5
26	0,64	0,39	0,85	0,26	8	0,7
27	0,64	0,40	0,86	0,25	9	0,9

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

Fonte: Elaborado pelo autor.

Os resultados demonstraram que o ajuste dos valores de SDC por um fator de 0,5 resultou em parâmetros estatísticos ligeiramente mais robustos (modelo 25). Este modelo foi selecionado para as análises subsequentes.

Os métodos de QSAR CoMFA e CoMSIA geram um número muito grande de descritores moleculares. Para alguns conjuntos de dados esta característica é

traduzida na utilização de uma quantidade excessiva de pontos na geração dos mapas de contorno. Este aspecto dificulta a interpretação da informação contida nestes mapas. Este fenômeno foi observado para os modelos CoMSIA, criando a necessidade de se eliminar descritores com valores muito baixos. Com este objetivo, a técnica de filtragem de colunas foi utilizada. Diferentes valores de corte foram empregados e mapas com melhor resolução foram obtidos aplicando-se um valor de corte de 2 kcal/mol. Após a determinação do valor de corte ótimo, foi avaliada a influência de alterações no número de componentes sobre os modelos. A redundância foi aferida por meio dos valores de dq^2/dr^2 . A tabela 17 mostra os parâmetros estatísticos obtidos de acordo com a variação do número de componentes.

Tabela 17 - Resultados CoMSIA obtidos em função do número de componentes (SDC = 0,5; alinhamento 1).

Modelo	q^2	SEP	r^2	SEE	N	dq^2/dr^2
28	0,61	0,40	0,81	0,29	5	0,78
29	0,63	0,40	0,82	0,28	6	0,81
30	0,64	0,40	0,83	0,27	7	0,97
31	0,63	0,40	0,84	0,26	8	1,07
32	0,63	0,42	0,86	0,25	9	1,14

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; N, Número de componentes; dq^2/dr^2 , variação de q^2 em função de r^2 .

Fonte: Elaborado pelo autor.

De acordo com a tabela 17, o modelo 30 apresentou maior consistência estatística. A estabilidade deste modelo foi avaliada pelo método LMO ($q^2 = 0,61$ e $0,62$ para cinco e dez grupos de validação, respectivamente). Em seguida, este modelo foi submetido ao processo de validação externa. A predição dos valores de atividade para o conjunto teste resultou em um valor de $r^2_{\text{pred}} = 0,79$. A tabela 18 mostra os valores experimentais e preditos de pIC_{50} para a série teste e a figura 19 ilustra a correlação entre estes valores para todo o conjunto de dados.

Tabela 18 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA – alinhamento 1).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo^a
2	4,30	4,69	-0,39
8	4,46	4,73	-0,27
13	4,49	4,52	-0,03
19	4,76	4,34	0,42
21	4,65	4,55	0,10
22	4,86	4,65	0,21
30	4,30	4,34	-0,04
34	4,94	4,55	0,39
38	4,30	4,64	-0,34
43	4,00	4,12	-0,12
52	5,22	4,89	0,33
64	4,00	4,07	-0,07
71	4,00	4,55	-0,55
74	4,00	4,24	-0,24
76	4,76	4,52	0,24
86	6,48	5,88	0,60
93	5,76	5,79	-0,03
95	5,85	5,68	0,17
97	4,85	4,62	0,23
102	5,60	6,07	-0,47
109	5,30	5,34	-0,04
116	4,00	4,07	-0,07
121	4,74	4,46	0,28
122	4,00	4,25	-0,25
126	4,74	4,08	0,66
127	5,63	5,99	-0,36

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

Fonte: Elaborado pelo autor.

Figura 19- Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMSIA (alinhamento 1).

Fonte: Elaborado pelo autor.

4.5.2 Modelos CoMSIA - Alinhamento 2

Estudos CoMSIA foram realizados com o alinhamento obtido através da superposição de farmacóforos. Os valores de α , SDC e filtragem de coluna (0,7; 0,5 e 2 kcal/mol, respectivamente) identificados nas análises CoMSIA anteriores, foram utilizados na modelagem para o alinhamento 2.

Inicialmente, foram produzidos modelos aplicando-se alterações no número de componentes. A redundância destes modelos foi verificada pela técnica de permutação dos valores da atividade biológica. Os resultados são exibidos na tabela 19.

Tabela 19 - Resultados CoMSIA obtidos em função da variação do número de componentes (SDC = 0,5; alinhamento 2).

Modelo	q^2	SEP	r^2	SEE	N	dq^2/dr^2
33	0,63	0,40	0,79	0,29	5	0,89
34	0,65	0,39	0,82	0,28	6	0,93
35	0,65	0,39	0,83	0,27	7	1,09
36	0,65	0,39	0,84	0,26	8	1,27
37	0,64	0,40	0,85	0,26	9	1,49

q^2 , Coeficiente de correlação com validação cruzada; r^2 , Coeficiente de correlação sem validação cruzada; SEP, Erro padrão de predição; SEE, Erro padrão de estimativa; N, Número de componentes; dq^2/dr^2 , variação de q^2 em função de r^2 .

Fonte: Elaborado pelo autor.

Como pode ser observado, os modelos gerados são muito semelhantes em relação aos principais parâmetros estatísticos. No entanto, o modelo 35 apresentou indicadores ligeiramente melhores. A estabilidade deste modelo foi avaliada pelo método LMO (obtidos $q^2 = 0,62$ e $0,65$, para cinco e dez grupos de validação cruzada, respectivamente). A predição dos valores de atividade para o conjunto teste resultou em um valor de $r^2_{\text{pred}} = 0,64$. A correlação entre os valores experimentais e preditos de pIC_{50} é apresentada na tabela 20. A figura 20 ilustra graficamente estes resultados.

Tabela 20 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA – alinhamento 2).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo ^a
2	4,30	4,63	-0,33
8	4,46	4,42	0,04
13	4,49	4,52	-0,03
19	4,76	3,97	0,79
21	4,65	4,50	0,15
22	4,86	4,66	0,20
30	4,30	3,68	0,62
34	4,94	4,44	0,50
38	4,30	4,47	-0,17
43	4,00	4,26	-0,26
52	5,22	4,94	0,28
64	4,00	4,13	-0,13
71	4,00	4,52	-0,52
74	4,00	4,62	-0,62

(Continua)

(Continuação)

Tabela 20 - Valores experimentais e preditos de pIC_{50} para o conjunto teste (modelo CoMSIA – alinhamento 2).

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo^a
76	4,76	4,91	-0,15
86	6,48	5,84	0,64
93	5,76	5,83	-0,07
95	5,85	5,70	0,15
97	4,85	4,73	0,12
102	5,60	5,52	0,08
109	5,30	5,04	0,26
116	4,00	4,41	-0,41
121	4,74	4,27	0,47
122	4,00	4,14	-0,14
126	4,74	4,79	-0,05
127	5,63	5,09	0,54

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

Fonte: Elaborado pelo autor.

Figura 20 - Valores experimentais e preditos de pIC_{50} para o melhor modelo CoMSIA (alinhamento 2).

Fonte: Elaborado pelo autor.

Os modelos CoMSIA obtidos para todos os alinhamentos demonstraram boa correlação entre os valores experimentais e preditos para a série teste. Estes resultados demonstraram que várias estratégias de alinhamento podem ser empregadas, sem que haja perda de consistência estatística. No entanto, a capacidade preditiva dos modelos foi influenciada de maneira significativa pelo alinhamento molecular. Novamente, o alinhamento baseado na sobreposição flexível da MCS gerou o modelo com maior capacidade preditiva ($r^2_{\text{pred}} = 0,79$).

4.6 Mapas de Contorno

Mapas de contorno 3D CoMFA e CoMSIA foram gerados para o conjunto de dados. A interpretação dos mapas é parte importante do processo de modelagem, porque as informações estruturais determinantes para a atividade biológica são assinaladas.

Nos mapas de contorno CoMFA, poliedros coloridos em verde representam locais em que grupos volumosos são favoráveis para o aumento da atividade biológica. Os mapas amarelos indicam regiões em que grupos volumosos são desfavoráveis para a atividade. Poliedros coloridos em azul e em vermelho representam, respectivamente, regiões em que grupos com carga positiva e negativa amplificam a atividade biológica.

A figura 21 mostra os mapas CoMFA para todas as estratégias de alinhamento utilizadas. O composto **85**, o mais potente da série ($pIC_{50} = 6,52$) é utilizado como referência. Poliedros verdes em torno do anel benzênico e do grupo etila ligado à posição *para* deste anel são encontrados em todos os modelos. Este resultado é consistente com os dados de atividade, considerando-se que o composto **86**, que também possui um grupo volumoso (metoxila) como substituinte nesta posição, apresenta um valor de $pIC_{50} = 6,48$. Em contrapartida, uma diminuição do volume dos grupos nesta região coincide com uma redução nos valores de atividade. Esta suposição é corroborada pela atividade dos compostos **85**, **88** e **91** (figura 22).

Também é possível observar um grande poliedro azul, próximo à posição 1 do anel benzênico, indicando que grupos com carga positiva podem ser explorados nesta região. O sistema π deste anel produz uma carga parcial negativa no centro do anel, e por consequência, uma carga parcial positiva na periferia. O perfil de atividade do conjunto de dados mostra que a ausência do sistema aromático nesta região reduz a potência dos compostos.

Figura 21 – Mapas de contorno CoMFA para o composto **85**. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2 e (C) Alinhamento 3.

Fonte: Elaborado pelo autor.

Figura 22 – Estrutura e atividade dos compostos **85**, **88**, e **91**, ilustrando a diminuição gradativa do volume dos grupos na região do anel aromático. Essa diminuição é acompanhada por uma redução nos valores de atividade.

Fonte: Elaborado pelo autor.

Na figura 23 são mostrados os mapas CoMSIA eletrostáticos e estereoquímicos para o composto **85**. Poliedros coloridos em verde e amarelo indicam locais em que grupos volumosos são, respectivamente, favoráveis e desfavoráveis para a atividade. Os mapas coloridos em azul representam locais em que grupos com carga positiva contribuem para o aumento da atividade. Os mapas em vermelho identificam locais em que grupos com carga negativa amplificam a atividade. Os modelos CoMSIA produziram mapas em verde na mesma posição contornada por poliedros verdes nos modelos CoMFA. Esta concordância entre os

mapas estereoquímicos produzidos por diferentes métodos ressaltam a importância de se explorar grupos volumosos nesta região. Os poliedros em azul apontam que grupos com carga positiva em torno do fragmento espaçador aumentam a atividade biológica dos compostos.

Figura 23 – Mapas de contorno CoMSIA eletrostáticos e estereoquímicos para o composto **85**. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2.

Fonte: Elaborado pelo autor.

A figura 24 ilustra os mapas hidrofóbicos, doadores e aceptores de interações de hidrogênio. Poliedros coloridos em cinza indicam que grupos hidrofóbicos são desfavoráveis para a atividade. Em contrapartida, poliedros amarelos indicam que grupos hidrofóbicos são favoráveis. Mapas em azul claro sugerem que doadores de interações de hidrogênio são favoráveis, ao passo que, poliedros em magenta preconizam que aceptores de interações de hidrogênio promovem a atividade. Poliedros vermelhos determinam que aceptores de interações de hidrogênio são prejudiciais para a atividade biológica.

Figura 24 – Mapas de contorno CoMSIA hidrofóbicos, doares e aceptores de interações de hidrogênio para o composto **85**. (A) Modelo gerado pelo alinhamento 1; (B) Alinhamento 2.

Fonte: Elaborado pelo autor.

A figura 24 exibe mapas em magenta próximos ao grupo ciano ligado ao anel piridínico, o que indica a importância de grupos aceptores de interações de hidrogênio nesta posição. De fato este grupo faz parte do fragmento estrutural essencial para a manutenção da atividade.

Nos mapas do alinhamento 1 (figura 24A), um poliedro cinza é observado na região da carbonila, indicando a importância de grupos polares. Outro poliedro cinza também é observado próximo ao substituinte *para* do anel benzênico, indicando que grupos hidrofóbicos são desfavoráveis dessa região. Nos mapas do alinhamento 2 (figura 24B), um poliedro magenta e um azul claro são observados próximos à carbonila, indicando que grupos doadores também são aceitáveis nesta região.

Conclusões e Perspectivas

5 Conclusões e Perspectivas

Nesta dissertação de mestrado, as relações quantitativas entre a estrutura e atividade para um conjunto de derivados piridínicos com atividade antichagásica foram exploradas. As moléculas investigadas possuem considerável atividade sobre a forma intracelular do *Trypanosoma cruzi*. Vale ressaltar que esta série de dados (estrutura e atividade) foi inteiramente produzida em trabalhos anteriores de nosso grupo de pesquisa em colaboração com o grupo liderado pelo Prof. Luiz Carlos Dias (IQ-UNICAMP), em um projeto que contou com o apoio da Organização Mundial da Saúde por meio de seu Centro de Referência Mundial em Química Medicinal para Doença de Chagas. Por essa razão, os compostos do conjunto de dados contam com características estruturais únicas, que são apropriadas para estudos de otimização molecular. A metodologia 2D de HQSAR e os métodos 3D CoMFA e CoMSIA foram empregados na construção dos modelos de QSAR. Informações estruturais importantes foram coletadas e correlacionadas com a atividade biológica. Dentre os principais resultados, destacam-se:

- Os modelos de QSAR gerados apresentam alta consistência interna e externa, bem como elevada estabilidade e baixa redundância, como demonstrado pelos indicadores estatísticos.
- O emprego de um conjunto teste confirmou a capacidade preditiva dos modelos para novos derivados piridínicos com atividade antichagásica.
- Diferentes métodos de alinhamento molecular com base em estratégias de LBDD foram capazes de produzir modelos de QSAR 3D estatisticamente robustos e relevantes.
- O método de alinhamento flexível com a sobreposição da máxima subestrutura comum produziu os melhores modelos CoMFA e CoMSIA.
- Os mapas de contribuição de HQSAR apontaram os principais fragmentos moleculares relacionados com a atividade biológica do conjunto de dados, indicando o núcleo piridínico como essencial à atividade biológica e

destacando a importância de substituintes do anel benzênico que possuam grupos aceptores de ligação de hidrogênio.

- Os mapas de contorno CoMFA e CoMSIA destacaram os principais aspectos de natureza 3D relacionados com a atividade dos compostos estudados.
- Os modelos 2D e 3D ressaltaram a importância do grupo ciano e dos substituintes ligados à posição *para* do anel benzênico. Estas informações serão úteis no planejamento, síntese e avaliação de novos agentes mais potentes contra o *T. cruzi*.
- Trabalhos futuros de nosso grupo no LQMC, particularmente na otimização molecular desta série de compostos, poderão considerar, entre outros, os aspectos destacados na figura 25: (i) a presença de substituintes volumosos na posição *para* do anel benzênico; (ii) a ausência de grupos volumosos na posição *orto* do anel benzênico; (iii) a substituição bioisostérica do anel benzênico por sistemas polares com características stereoquímicas semelhantes; e, (iv) a introdução de grupos positivos ou doadores de interações de hidrogênio na posição *orto* do anel benzênico.

Figura 25 – Estratégias a serem exploradas em trabalhos de otimização molecular para esta série de compostos.

Referências

REFERÊNCIAS

- 1 WORLD HEALTH ORGANIZATION. **Why are some tropical diseases called “neglected.”**Disponível em: <<http://www.who.int/features/qa/58/en/>> Acesso em: 11 jan. 2015
- 2 WORLD HEALTH ORGANIZATION. **Sustaining the drive to overcome the global impact of neglected tropical diseases.** Geneva: 2013. 138 p.
- 3 WORLD HEALTH ORGANIZATION. **Accelerating work to overcome the global impact of neglected tropical diseases.** Geneva: 2012. 14 p.
- 4 FENWICK, A. The global burden of neglected tropical diseases. **Public Health**, v. 126, n. 3, p. 233–236, 2012.
- 5 RIDLEY, R. G.; FLETCHER, E. R.; DOW, M. Making a difference : 30 years of TDR. **Nature Reviews: microbiology** v. 6, n. 5, p. 401–407, 2008.
- 6 SCHMUNIS, G. A; YADON, Z. E. Chagas disease: a new worldwide challenge. **Nature Outlook: chagas disease**, v. 465, n. 7301, p. 14–21, 2010.
- 7 EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, n. 5733, p. 409–415, 2005.
- 8 ATWOOD, J. A et al. The *Trypanosoma cruzi* proteome. **Science**, v. 309, n. 5733, p. 473–476, 2005.
- 9 COURA, J. R.; CASTRO, S. L. DE. A critical review on chagas disease chemotherapy. **Memórias do Instituto Oswaldo Cruz**, v. 97, n.1, p. 3–24, 2002.
- 10 DIAS, L. C. et al. Quimioterapia da doença de chagas: estado da arte e perspectivas no desenvolvimento de novos fármacos.**Química Nova**, v. 32, n. 9, p. 2444–2457, 2009.
- 11 CLAYTON, J. Chagas disease 101. **Nature Outlook: chagas disease**, v. 465, n. 7301, p. 4–5, 2010.
- 12 SCHOFIELD, C. J.; GORLA, D. E. (Comp.) Who , how , what and where ? **Nature Outlook: chagas disease**, v. 465, n. 7301, p. 4–5, 2010.
- 13 RASSI Jr, A.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **Lancet**, v. 375, n. 9723, p. 1388–1402, 2010.
- 14 WERMUTH, C. G. **The practice of medicinal chemistry.** 3rd. ed. London: Academic Press, 2008. p.63.

15 GUIDO, R. V. C.; ANDRICOPULO, A. D.; OLIVA, G. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estudos Avançados**, v. 24, n. 70, p. 81–98, 2010.

16 CAŁA, J. et al. Cloud computing for fast prediction of chemical activity. **Future Generation Computer Systems**, v. 29, n. 7, p. 1860–1869, 2013.

17 BASKIN, I.; VARNEK, **Chemoinformatics approaches to virtual screening**. Cambridge: Royal Society of Chemistry, 2008. p. 338.

18 XIANG, M. et al. Computer-Aided drug design: lead discovery and optimization. **Combinatorial Chemistry & High Throughput Screening**, v. 15, n. 4, p. 328–337, 2012.

19 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.

20 CHERKASOV, A. et al. QSAR Modeling: where have you been? where are you going to? **Journal of Medicinal Chemistry**, v. 57, n. 12, p. 4977–5010, 2013.

21 VERMA, J.; KHEDKAR, V.; COUTINHO, E. 3D-QSAR in drug design - a review. **Current Topics in Medicinal Chemistry**, v. 10, n. 1, p. 95–115, 2010.

22 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.

23 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.

24 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.

25 GRAMATICA, P. Principles of QSAR models validation: internal and external. **QSAR & Combinatorial Science**, v. 26, n. 5, p. 694–701, 2007.

26 TROPSHA, A. Best practices for QSAR model development, validation, and exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, 2010.

27 RU, C.; RU, G. γ -Randomization and its variants in QSPR / QSAR. **Journal of Chemical Information and Modeling**, v. 47, n. 6, p. 2345–2357, 2007.

- 28 CLARK, R. D.; FOX, P. C. Statistical variation in progressive scrambling. **Journal of Computer-Aided Molecular Design**, v. 18, n. 7-9, p. 563–576, 2004.
- 29 LOWIS, D. R. HQSAR: a new , highly predictive QSAR technique. **Tripos Technical Notes** v. 1, n. 5, p. 1–17, 1997.
- 30 SALUM, L. B.; DIAS, L. C.; ANDRICOPULO, A. D. Fragment-based QSAR and molecular modeling studies on a series of discodermolide analogs as microtubule-stabilizing anticancer agents. **QSAR & Combinatorial Science**, v. 28, n. 3, p. 325–337, 2009.
- 31 WOLD, S.; SJOSTROM, M. PLS-regression : a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.
- 32 CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins. **Journal Of The American Chemical Society**, v. 110, n. 11, p. 5959–5967, 1988.
- 33 KLEBE, G. Comparative molecular similarity indices analysis: CoMSIA. **Perspectives in Drug Discovery and Design**, v. 12, p. 87–104, 1998.
- 34 RONCAGLIONI, A.; BENFENATI, E. In silico-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects. **Chemical Society Reviews**, v. 37, n. 3, p. 441–50, 2008.
- 35 ROY, K. K.; DIXIT, A.; SAXENA, A. K. An investigation of structurally diverse carbamates for acetylcholinesterase (AChE) inhibition using 3D-QSAR analysis. **Journal of Molecular Graphics & Modelling**, v. 27, n. 2, p. 197–208, 2008.
- 36 KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature Reviews: drug discovery**, v. 3, n. 11, p. 935–49, 2004.
- 37 WOLBER, G. et al. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. **Drug Discovery Today**, v. 13, n. 1-2, p. 23–9, 2008.
- 38 YANG, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug Discovery Today**, v. 15, n. 11-12, p. 444–50, 2010.
- 39 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. 773–88, 2006.
- 40 PATEL, Y. et al. A comparison of the pharmacophore identification programs : Catalyst , DISCO and GASP. **Journal of Computer-Aided Molecular Design**, v. 16, n 8, p. 653–681, 2002.

- 41 RICHMOND, N. J. et al. GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. **Journal of Computer-Aided Molecular Design**, v. 20, n. 9, p. 567–587, 2006.
- 42 FERREIRA, R. S. et al. Synthesis, biological evaluation, and structure-activity relationships of potent noncovalent and nonpeptidic cruzain inhibitors as anti-trypanosoma cruzi agents. **Journal of Medicinal Chemistry**, v. 57, n. 6, p. 2380-2392, 2014.
- 43 DIAS, Luis C. et al. Quimioterapia da doença de Chagas: estado da arte e perspectivas no desenvolvimernto de novos fármacos. **Química Nova**, v. 32, n. 9, p. 2444-2457, 2009.
- 44 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.
- 45 SILVA, D. C. et al. Identification of structural and electronic features for a series of MCH1R antagonists. **Medicinal Chemistry: hilversum**, v. 9, n. 1, p. 22-31, 2013.
- 46 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
- 47 GUIDO, R. V. C. et al. Classical and hologram QSAR studies on a series of inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. **QSAR & Combinatorial Science**, v. 27, n. 6, p. 768-781, 2008.