



UNIVERSIDADE FEDERAL DA BAHIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FARMÁCIA

FRANCIELLE MARTINS DE MELO

**Estudo *in silico* de inibidores da lanosterol 14 alfa-
desmetilase de *Trypanosoma cruzi***

SALVADOR
2012

FRANCIELLE MARTINS DE MELO

Estudo *in silico* de inibidores da lanosterol 14 alfa-desmetilase de *Trypanosoma cruzi*

Dissertação apresentada ao Programa de Pós-Graduação em Farmácia da Universidade Federal da Bahia, como requisito final para obtenção do título de Mestre em Ciências Farmacêuticas.

Orientador: Prof. Dr. Marcelo Santos Castilho

SALVADOR
2012

Sistema de Bibliotecas da UFBA

Melo, Francielle Martins de
Estudo *in silico* de inibidores de lanosterol 14alfa-desmetilase de *Trypanosoma cruzi*
Francielle Martins de Melo. - 2012.
127 f. : il.

Orientador: Prof. Dr. Marcelo Santos Castilho.
Dissertação (mestrado) - Universidade Federal da Bahia, Faculdade de Farmácia, Salvador,
2012.

1. Chagas, Doença de - Tratamento. 2. Enzimas - Inibição. 3. *Trypanosoma cruzi*.
4. Drogas - Resistência em microorganismos. 5. Medicamentos - Utilização. I. Castilho, Marcelo
Santos. II. Universidade Federal da Bahia. Faculdade de Farmácia. III. Título.

CDD - 616.9363
CDU - 616.937



UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE FARMÁCIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FARMÁCIA



TERMO DE APROVAÇÃO

FRANCIELLE MARTINS DE MELO

ESTUDOS *IN SILICO* DE INIBIDORES DE LANOSTEROL 14ALFA- DESMETILASE DE *TRYPANOSOMA CRUZI*

Dissertação submetida ao Programa de Pós-Graduação em Farmácia (nível Mestrado Acadêmico) da Faculdade de Farmácia da Universidade Federal da Bahia, como requisito parcial para a obtenção do título de Mestre em Farmácia.

Aprovada em 18 de maio de 2012.

BANCA EXAMINADORA

Dr. Marcelo Santos Castilho
Universidade Federal da Bahia
Orientador

Dr.ª. Carolina Horta Andrade
Universidade Federal de Goiás

Dr. Gustavo Henrique Goulart Trossini
Universidade de São Paulo

AGRADECIMENTOS

A Deus, aos meus pais, Altair e Maria Verônica, aos meus irmãos Márcio, Marcelo e Flavielle, familiares e amigos pelo apoio, incentivo, paciência, carinho e companheirismo diário.

Ao meu orientador Prof. Dr. Marcelo Castilho pela oportunidade, excelente orientação, discussões e compromisso.

Aos colegas do LaBiMM pelo incentivo, discussões e brincadeiras

Aos professores e colegas da Pós Graduação em Farmácia

Ao CNPQ pelo apoio financeiro

Apoio Financeiro

CNPq (Bolsista nº PROC 552748/2010-2)

Os recursos financeiros necessários à realização do trabalho experimental obtidos do Projeto CNPq 552748/2010-2 “Estudo *in silico* de inibidores da lanosterol 14alfa-desmetilase de *Trypanosoma cruzi*” em consolidação ao Programa de Pós-Graduação em Farmácia (mestrado) da Universidade Federal da Bahia.

A alegria está na luta, na tentativa, no sofrimento envolvido. Não na vitória propriamente dita.

Mahatma Gandhi

RESUMO

A doença de Chagas é uma doença negligenciada de grande impacto sobre a população da América Latina. No Brasil estima-se que aproximadamente 50.000 novos casos de doença de Chagas ocorram anualmente e que cerca de 2.500.000 indivíduos estejam infectados. Apesar disso, o arsenal terapêutico disponível para o tratamento da doença (nifurtimox e benzonidazol) apresenta pouca ação sobre a fase crônica e sérios efeitos colaterais. Além disso já foi relatado o aparecimento de cepas resistentes aos medicamentos disponíveis. Diante desses fatos, torna-se importante a identificação de novos protótipos a fármacos ativos contra a fase crônica da doença. Sabe-se que antifúngicos azólicos (posaconazol, ravuconazol, itraconazol e cetoconazol) inibem a enzima lanosterol 14 alfa-desmetilase do *Trypanosoma cruzi* e por essa razão impedem a síntese do ergosterol, o principal esterol de membrana do parasita. Moléculas com esse perfil de atividade são ativas contra a forma amastigota do parasito, apresentando resultados positivos no tratamento da fase crônica da doença de Chagas. Esses dados sugerem que o desenvolvimento de novos inibidores da via do ergosterol é uma alternativa interessante para o tratamento da doença de Chagas. Visando contribuir para esse objetivo foram desenvolvidos modelos de QSAR 2D e 3D para um conjunto de 155 moléculas ativas contra *T. cruzi*. Modelos de QSAR 2D, baseado, em descritores bidimensionais ($r^2= 0,86$, $q^2=0,9$, $r^2_{pred}=0,84$, com 4PCs) e Holograma moleculares ($r^2= 0,85$, $q^2=0,91$, $r^2_{pred}=0,82$, com 6PCs) apresentam parâmetros estatísticos satisfatórios e sugerem que propriedades estéreas da amina ligada a anel aromático estão relacionadas com aumento na atividade dos compostos, enquanto que ramificações estruturais tem o efeito contrário. Visando complementar o estudo das exigências estéreas e eletrônicas que determinam a atividade biológica dos derivados azólicos estudados, foram desenvolvidos modelos de QSAR 3D pela técnica de análise comparativa de campos moleculares (CoMFA). O modelo do alinhamento realizado pelo acoplamento molecular propiciou poder preditivo satisfatório ($q^2 = 0,74$, $r^2 = 0,92$, $r^2_{pred} = 0,70$, com 5 PCs). O modelo realizado pelo alinhamento por similaridade morfológica dos 5 ligantes de maior atividade proporcionou maior poder preditivo ($q^2= 0,75$, $r^2=0,91$, $r^2_{pred}=0,77$, com 6PCs). A análise dos mapas de contorno sugere que a amina primária ligada ao anel aromático tem contribuição estérea e eletrônica importante, assim como substituintes na posição meta do anel *p*-anilina. O anel fenil na posição meta, assim como o anel imidazol podem diminuir a atividade dos compostos. As informações obtidas através dos modelos quimiométricos descritos acima são de suma importância para guiar o planejamento de moléculas mais potentes contra *T. cruzi*.

Palavras-chave – Doença de Chagas, Trypanosoma cruzi, Lanosterol 14 alfa-desmetilase, QSAR 2D, QSAR 3D.

ABSTRACT

Chagas' disease has a major impact over the population of Latin America. This neglected disease affects 2,500,000 patients and approximately new 50,000 cases are identified annually. Despite that, there are few drugs available to treat this ailment (nifurtimox and benznidazole), which are not effective against the chronic phase and show many side-effects. In addition, drug-resistant strains have been reported. These facts underscore the importance of discovering lead compounds that are active against the chronic phase of Chagas' disease. Antifungal compounds such as posaconazole, revuconazole and cetoconazole also inhibit the lanosterol 14- α demethylase from *Trypanosoma cruzi*, thus blocking the ergosterol pathway. Moreover, such compounds are active against amastigote *T. cruzi* and show promising results for chronic Chagas' disease treatment. Therefore, developing new ergosterol biosynthesis inhibitors has been considered as a good strategy to improve Chagas' disease treatment. Aiming at this goal, 2D and 3D QSAR models were built for a dataset of 155 compounds that had been assayed against *T. cruzi*. 2D-QSAR models built with topological descriptors ($r^2=0.86$, $q^2=0.9$, $r^2_{\text{pred}}=0.84$, with 4PCs) or molecular hologram ($r^2=0.85$, $q^2=0.91$, $r^2_{\text{pred}}=0.82$, with 6PCs) show reasonable statistics parameters and suggest that steric properties from the NH_2 bound to phenyl ring increase potency, whereas moieties branching has the opposite effect. In order to further investigate the steric and electronic requirements for biological activity, 3D-QSAR models were developed by Comparative Molecular Field Analysis (CoMFA). The model of alignment by docking presented good predictive model ($q^2=0.74$, $r^2=0.92$, $r^2_{\text{pred}}=0.70$, com 5 PCs). The model of alignment by morphological similarity towards the 5 most activity compounds presented most predictive model ($r^2=0.75$, $q^2=0.91$, $r^2_{\text{pred}}=0.77$, with 6PCs). Contour map analysis not only supports the hypothesis that NH_2 bound to phenyl has a positive steric or electronic contribution to activity but also meta substitute on *p*-aniline. The ring phenil in the position meta as well as the ring imidazole would reduce the activity. The information gathered from all these models shall be useful to guide the development of second generation molecules with increased potency against *T. cruzi*.

Palavras-chave – Chagas disease, *Trypanosoma cruzi*, Lanosterol 14 alpha-desmethylase, QSAR 2D, QSAR 3D.

LISTA DE FIGURAS

- Figura 1.** Relação entre os investimentos em Pesquisa e Desenvolvimento feitos pela indústria farmacêutica e fabricantes da América. Em função do número de novas entidades químicas (NCE) e imunobiológicos. 18
- Figura 2.** Exemplo de processos baseados no ligante. No modelo farmacofórico, um conjunto de características estéreas e eletrônicas em comum de um conjunto de dados são selecionadas por serem importantes na interação com o alvo macromolecular. No modelo por similaridade química, representado pela similaridade morfológica do programa surflexsim, cria-se pontos de observação na molécula referência de acordo com as características de Van der Waals e ligações de hidrogênio. Um conjunto de pontos externos (triângulo) são formados, com a área representando as características moleculares e comparados com outros triângulos em busca de características moleculares semelhantes. 23
- Figura 3.** Descritores moleculares utilizados no desenvolvimento de modelos de QSAR. Representado em A descritores topológicos, constitucionais, fisicoquímicos etc, usados em modelos como QSAR baseado em descritores bidimensionais. Em B, descritores fragmentais, usados, por exemplo, na metodologia de HQSAR com a utilização de hologramas. 24
- Figura 4.** Representação da metodologia do acoplamento molecular. Interações entre o ligante e os aminoácidos presente no sítio ativo da Lanosterol 14 alfa-desmetilase são analisados. 25
- Figura 5.** Etapas envolvidas na geração do modelo de CoMFA. 26
- Figura 6.** Representação do número de doenças negligenciadas por região do globo terrestre 27
- Figura 7.** Diversidade de espécies de insetos triatomíneos (linhagem Rhodniini) por unidade taxonômica operacional (OTU, do inglês *operational taxonomic units*) na América Latina. 32
- Figura 8.** Rota de migração da América Latina e estimativa do número total de indivíduos infectados em países não endêmicos. 33
- Figura 9.** Número de casos de pessoas infectadas pelo *Trypanosoma cruzi* nos anos de 2000 a 2010 no Brasil. Observa-se que, dentre os tipos de contaminação, a transmissão por via oral é a predominante. 34
- Figura 10.** Provável mecanismo de ação dos fármacos benzonidazol e nifurtimox. 35
- Figura 11.** Via biossintética resumida do ergosterol. 39

Figura 12. Sítio ativo e bolsão hidrofóbico da enzima lanosterol 14 alfa-desmetilase de <i>T. cruzi</i> com o inibidor posaconazol. O sitio ativo está representado pela hélice B'/alça B'C, de coloração azul claro, onde situa-se os aminoácidos tirosina 103, metionina 106, fenilalanina 110 e tirosina 116; pela hélice C, de coloração azul claro e situada atrás da figura (não visível), que contém o aminoácido leucina 127; hélice I, de coloração amarela, onde estão presentes os aminoácidos alanina 297, fenilalanina 290, alanina 291 e tirosina 295 e uma alça representada em laranja, com a presença dos aminoácidos leucina 356, leucina 357 e metionina 360. Pode observar a presença de um túnel hidrofóbico, com a hélice F" em verde, com os aminoácidos prolina 210, alanina 211, valina 213 e fenilalanina 214 e a alça em vermelho, com os aminoácidos tirosina 457, histidina 458, tirosina 459 e metionina 460	40
Figura 13. Fármacos que apresentam atividade contra a enzima lanosterol 14 alfa-desmetilase de <i>T. cruzi</i> <i>in vitro</i> .	41
Figura 14. Distribuição dos derivados azólicos no espaço químico definido por PC1 e PC2. Para a divisão do grupo teste optou-se por escolher moléculas que fazem parte de todos os quadrantes.	66
Figura 15. Distribuição dos derivados azólicos no grupo treino (em preto) e teste (em vermelho), segundo sua atividade biológica.	66
Figura 16. Valores reais versus valores preditos de pED ₅₀ de derivados imidazólicos inibidores da enzima lanosterol 14 alfa-desmetilase de <i>T. cruzi</i> no modelo de RLM.	69
Figura 17. Gráfico de pontuação do modelo de PCA para os derivados azólicos ativos contra a enzima Lanosterol 14 alfa-desmetilase de <i>T. cruzi</i> .	70
Figura 18. Gráfico de pesos do modelo de PCA demonstrando os descritores mais importantes para PC2.	71
Figura 19. Valores reais versus valores preditos do modelo de PLS baseado em descritores bidimensionais com os 53 descritores oriundos do programa Mobydigs 1.0.	72
Figura 20. Valores experimentais e preditos de pED ₅₀ de acordo com o modelo de PLS com 20 descritores bidimensionais.	73
Figura 21. Vetor de regressão do modelo de QSAR baseado em descritores bidimensionais.	74
Figura 22. Importância para o descritor fragmental nArNH ₂ e TI ₂ para a atividade biológica de acordo com o modelo de QSAR. Os gráficos 1 e 2 demonstram a relação da variância dos descritores de acordo com pED ₅₀ das moléculas. É observável que grupos de moléculas com atividade biológica baixa possuem baixa variância do descritor, enquanto algumas moléculas de com atividade biológica alta contêm alta variância do	76

descriptor. Moléculas com alto valor de $nArNH_2$ (gráfico 1 dessa figura, A, nº110 da Tabela 2, com pED_{50} de 9,45, com duas aminas primárias ligadas ao anel aromático) tem alta atividade biológica, enquanto que a molécula B (nº33 da Tabela 2, com pED_{50} de 7,61, com 1 amina primária ligada ao anel aromático) tem valor de atividade biológica menor e a molécula C (nº36 da Tabela 2, com pED_{50} de 6,48, sem aminas primárias) tem baixo valor de atividade. O mesmo raciocínio para o gráfico 2 dessa figura. Moléculas com baixo valor de TI2 (D, nº 83 da Tabela 2) geralmente possuem atividade biológica reduzida, com pED_{50} de 6,00. Moléculas com alto valor de TI2 (E, nº34 da Tabela 2; F, nº106 da Tabela 2) possuem atividade biológica aumentada, com pED_{50} de 8,23 e 8,95 respectivamente.

Figura 23. Relação entre a atividade biológica dos inibidores e a variância dos descritores BLI e BAC. Observa-se que moléculas com atividade biológica baixa apresentam elevado valor desse descritor, enquanto que moléculas de atividade biológica alta, esse valor decresce. No gráfico 3 dessa figura a molécula A de nº 31 da Tabela 2 com pED_{50} de 5,00, apresenta várias ligações com átomos passíveis de fazer ressonância e tem valor elevado de BLI. As moléculas B e C, representada pela molécula 108 e 109 da Tabela 2, com pED_{50} 9,34 e 9,35 respectivamente, não possui tantas ligações ressonantes como a molécula A. No gráfico 4 dessa figura, a molécula D está representada uma molécula ramificada (nº 4 da Tabela 2, pED_{50} = 6,52), enquanto que as demais possuem ramificações menores (E nº 61; pED_{50} = 8,00 e F, nº82 da Tabela 2, pED_{50} = 7,69).

77

Figura 24. Importância dos descritores para as PCs segundo modelo de PLS. Considerando PC3, os descritores $nArNH_2$ e BAC contribuem para a separação dos compostos, enquanto nH tem baixa importância no modelo.

78

Figura 25. Valores reais versus valores preditos de pED_{50} de derivados imidazólicos inibidores da enzima lanosterol 14 alfa-desmetilase de *T. cruzi* no modelo de HQSAR.

81

Figura 26 Mapa de contribuição para uma molécula com alta atividade pED_{50} = 9,45 (A, nº 110, Tabela 2), uma molécula de atividade intermediária pED_{50} = 7,69 (B, nº 82, Tabela 2) e uma molécula de baixa potência pED_{50} = 5,00 (C, nº 20, Tabela 2 de acordo com o melhor modelo de HQSAR (A/B/C/H/Ch/DA).

83

Figura 27. Interpretação dos modelos de HQSAR e QSAR de descritores bidimensionais. Em A está demonstrado em negrito o descritor $nArNH_2$, correspondente ao mapa de contribuição. Em B, em negrito, pode representar o descritor TI2 e de acordo com parte do mapa de contribuição. Em C, em negrito, pode representar o descritor BAC e BLI, como de acordo com o mapa do modelo de HQSAR.

84

Figura 28. Alinhamento dos derivados imidazólicos pelo método da máxima estrutura comum (B). Os grupos químicos considerados comuns a todas as moléculas estão representados em vermelho da molécula A.

86

Figura 29. Comparação entre o posaconazol cristalográfico (em laranja) e a solução proposta com o auxílio do programa GOLD 4.1.2 (em verde). Nota-se que a orientação do anel imidazol da solução não interage com o ferro do grupo heme.	90
Figura 30. Alinhamento dos compostos do grupo treinamento através do acoplamento molecular sem restrição. Nota-se que alguns anéis imidazóis estão em diferentes locais no espaço (alguns estão inseridos na região circundada em vermelho).	91
Figura 31. Interação entre posaconazol e o ferro do grupo heme da enzima lanosterol 14- α desmetilase de <i>T. cruzi</i> (código pdb 3K1O, resolução 2,89).	93
Figura 32. Comparação entre o posaconazol cristalográfico (em laranja) e a solução proposta com o auxílio do programa Gold (em verde). Nota-se que a orientação da solução é semelhante a orientação do ligante cristalográfico.	93
Figura 33. Alinhamento dos derivados imidazólicos pelo método GOLD. No círculo em vermelho, localiza-se a interação do anel imidazol com o ferro do grupo heme.	94
Figura 34. Valores reais vs valores preditos dos inibidores imidazólicos obtidos pelo método de acoplamento molecular.	97
Figura 35. Mapa de contorno do ligante (vermelho). Os campos estereoquímicos representado em verde e amarelo devido o cálculo de Lennard-Jonnes da interação entre a carga de prova e os átomos dos ligantes. Em A está representado um ligante com alta atividade ($pED_{50} = 9,45$, número 110 da Tabela 2); em B, o ligante de menor atividade ($pED_{50} = 5,00$, número 20 da Tabela 2).	99
Figura 36. Mapa de contorno do ligante (laranja). Os campos eletrostáticos representado em vermelho e azul devido o cálculo de Coulomb da interação entre a carga de prova e os átomos dos ligantes. Em A está representado um ligante de alta atividade biológica ($pED_{50} = 9,45$, número 110 da Tabela 2); em B ligante de baixa atividade ($pED_{50} = 5,00$, número 20 da Tabela 2).	100
Figura 37. Ligantes cristalizados que interagem com a enzima Lanosterol 14 α -desmetilase de <i>T. cruzi</i> . Em A, o ligante 4-(4-clorofenil)-N-[2-(1H-imidazol-1-il)-1-feniletil] benzamida (VNF), que interage com uma parte específica do sítio no qual nenhum outro ligante interage (héliceB'/loopB'C; hélice C e porção N terminal da hélice I). Em B, a molécula do fluconazol e em C a do posaconazol.	102
Figura 38. Alinhamento dos derivados imidazólicos sobre o posaconazol.	103
Figura 39. Modelo de similaridade das 5 moléculas mais potentes do conjunto de dados.	104

- Figura 40.** Alinhamento dos derivados imidazólicos pelo método surflex-sim. Representação em A do conjunto treinamento e em B do conjunto teste. Em círculo vermelho, a localização dos anéis imidazol. 105
- Figura 41.** Relação das pontuações das conformações geradas com o auxílio do programa Surfex-sim versus pED_{50} dos compostos. 105
- Figura 42.** Gráfico real versus predito do modelo de CoMFA com alinhamento do Surfex-sim. 108
- Figura 43.** Mapa de contorno estereoquímico dos derivados do modelo de CoMFA do alinhamento do Surfex-sim. (A) Molécula com alta atividade biológica, com pED_{50} de 9,45; (B) Molécula com baixa atividade, com pED_{50} de 5,00. 110
- Figura 44.** Mapa de contorno eletrostático dos derivados do modelo de CoMFA do alinhamento do Surfex-sim. (A) Molécula com atividade alta, com pED_{50} de 9,45 e (B) Molécula com baixa atividade, com pED_{50} de 5,00. 110

LISTA DE TABELAS

Tabela 1. Fármacos e imunobiológicos aprovados para doenças negligenciadas no período de 2000 a 2009.	29
Tabela 2. Moléculas do grupo treinamento.	43
Tabela 3. Moléculas do grupo teste.	53
Tabela 4. Valores da variância explicada e acumulada nas PCs do modelo de PCA.	65
Tabela 5. Valores reais, predito e residuais para os compostos do grupo teste de acordo com o modelo de PLS.	73
Tabela 6. Descritores que contribuem para a atividade dos compostos no modelo de regressão pelo método dos mínimos quadrados parciais.	74
Tabela 7. Influência da distinção de fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com tamanho de fragmento padrão (4 – 7).	80
Tabela 8. Influência do tamanho dos fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com distinção de fragmento ABCHChDA.	80
Tabela 9. Influência do tamanho dos fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com distinção de fragmento ABCHChDA.	81
Tabela 10. Valores reais, preditos e residuais para o conjunto teste do modelo A/B/C/H/Ch/DA de tamanho de fragmento 4-7.	82
Tabela 11. Resultados do alinhamento de máxima estrutura comum.	87
Tabela 12. Valores dos parâmetros estatísticos do alinhamento realizado com o auxílio do programa Gold 4.1.2	91
Tabela 13. Resultados do CoMFA obtidos com alinhamento do programa GOLD.	95
Tabela 14. Parâmetros estatísticos para análise as estabilidade do modelo do alinhamento baseado na estrutura com 10 grupos escolhidos aleatoriamente.	96
Tabela 15. Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA ccom o alinhamento do GOLD 4.1.2.	96
Tabela 16. Resultados do CoMFA com alinhamento do Posaconazol pela metodologia de similaridade morfológica realizado com o auxílio do	103

programa Surfex-sim.

Tabela 17. Resultados estatísticos dos modelos CoMFA obtidos a partir do alinhamento por similaridade química utilizando os 5 inibidores mais potentes. 106

Tabela 18. Valores de r^2_{boot} e SEE_{boot} para avaliar a estabilidade do modelo de similaridade morfológica do modelo dos cinco ligantes mais potentes. 107

Tabela 19. Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA com o alinhamento do Surfex-sim. 107

LISTA DE SIGLAS E ABREVIATURAS

2D – Bidimensional

3D - Tridimensional

A - Átomo

B – Ligação

C – Conectividade

Ch – Quiralidade

CoMFA - Análise Comparativa dos Campos Moleculares (do inglês *Comparative Molecular Field Analysis*)

DA – Doador e aceptor de hidrogênio

DALY - Anos de Vida Perdidos (do inglês *Disability Adjusted Life Years*)

DNA – Ácido Desoxirribonucleico

E – Campo eletrostático

ED₅₀ – Dose efetiva

GOLD – Genetic Optimization for Ligand Docking

H – Hidrogênio

HGPRT - Hipoxantina-Guanina Fosforibosil Transferase

HQSAR – Holograma QSAR

HTS – Ensaio em Larga Escala (do inglês *high throughput screening*)

IOWH - Institute for One World Health

LBDD – Planejamento de Fármacos Baseado no Ligante (do inglês *ligand-based drug design*)

LMO – Deixe vários para fora (do inglês *Leave-Many-Out*)

LOO – Deixe um para fora (do inglês *Leave-One-Out*)

MMV - *Medicines for Malaria Venture*

MLR – Regressão Linear Múltipla (do inglês *Multiple Regression Linear*)

NCE – Nova Entidade Química (do inglês *New Chemical Entities*).

PC – Componente Principal (do inglês *Principal Components*)

PCA – Análise de Componentes Principais (do inglês *Principal Components Analysis*)

PDB – Banco de Dados de Proteínas (do inglês *Protein Data Bank*)

PLS - Mínimos Quadrados Parciais (do inglês *Partial Least Square*)

PPP – Parcerias Público-Privada

QSAR – Relação Estrutura-Atividade Quantitativa (do inglês *Quantitative Structure-Activity Relationship*).

RMSD – Desvio da Raiz Média Quadrada (do inglês *Root-Mean-Square Deviation*)

RNA – Ácido Ribonucleico

S – Campo estereoquímico

SAR- Relação Estrutura-Atividade (do inglês *Structure-Activity Relationship*)

SBDD – Planejamento de Fármacos Baseado no Ligante (do inglês *Structure-Based Drug Design*).

SDC - Coeficiente de Desvio Padrão (do inglês *Standard Deviation Coefficient*)

SEE - Erro Padrão da Estimativa (do inglês *Standard Error Estimate*)

SEP - Erro Padrão de Predição (do inglês *Standard Error Prediction*)

VNF - 4-(4-clorofenil)-N-[2-(1H-imidazol-1-il)-1-feniletil] benzamida

VS – Triagem Virtual (do inglês *virtual Screening*)

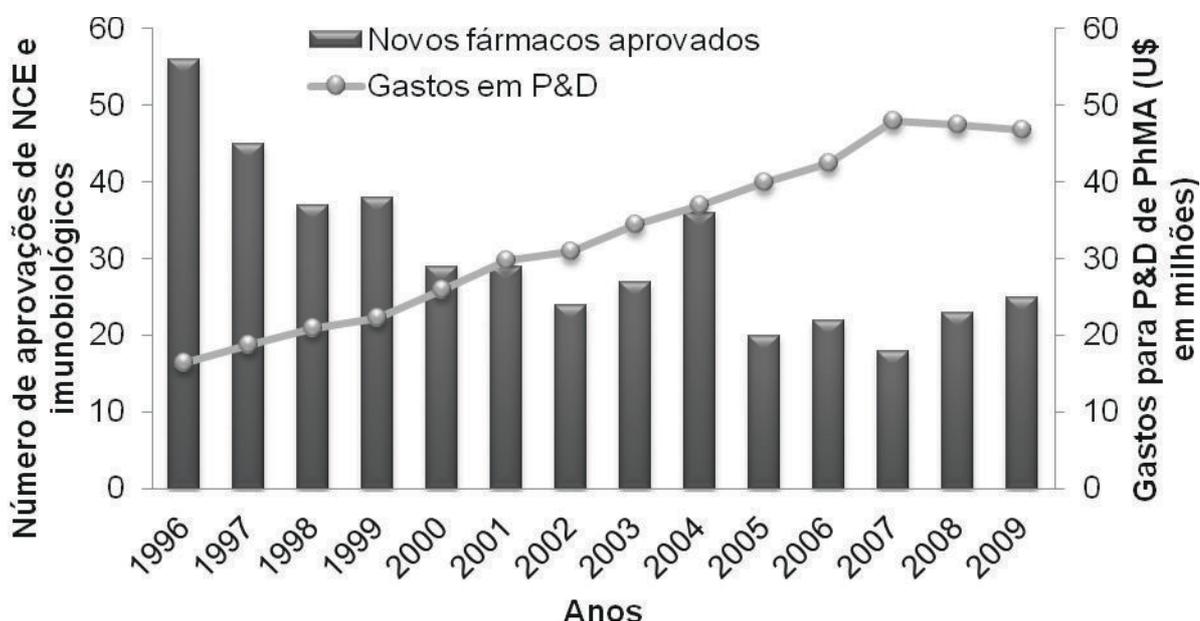
SUMÁRIO

1. INTRODUÇÃO	18
1.1 PROCESSO DE IDENTIFICAÇÃO E OTIMIZAÇÃO DE MOLÉCULAS BIOATIVAS	19
1.2 FERRAMENTAS COMPUTACIONAIS NO DESENVOLVIMENTO DE FÁRMACOS	21
1.3 DOENÇAS TROPICAIS NEGLIGENCIADAS	27
1.3.1 DOENÇA DE CHAGAS	30
1.3.1.1 DESENVOLVIMENTO DE FÁRMACOS CONTRA DOENÇAS DE CHAGAS	35
1.3.1.2 INIBIDORES DA VIA DE ERGOSTEROL	37
2. OBJETIVOS	42
2.1 OBJETIVO GERAL	42
2.2 OBJETIVOS ESPECÍFICOS	42
3. MATERIAL E MÉTODOS	43
3.1 CONJUNTO DE DADOS	43
3.2 MÉTODOS QUIMIOMÉTRICOS	56
3.2.1 CÁLCULO E SELEÇÃO DOS DESCRITORES	56
3.2.2 ANÁLISE DE COMPONENTES PRINCIPAIS	56
3.2.3 QSAR BASEADO EM DESCRITORES BIDIMENSIONAIS	57
3.2.4 HOLOGRAMA QSAR (HQSAR)	58
3.2.5 ANÁLISE COMPARATIVA DOS CAMPOS MOLECULARES (COMFA)	58
3.2.5.1 ALINHAMENTO POR ACOPLAMENTO MOLECULAR	58
3.2.5.2 ALINHAMENTO POR SIMILARIDADE MOLECULAR	59
3.2.5.3 ALINHAMENTO PELA MÁXIMA SUBESTRUTURA	60
3.2.6 VALIDAÇÃO DOS MODELOS	61
3.2.6 VALIDAÇÃO DOS MODELOS	61
4. RESULTADOS E DISCUSSÃO	63
5. CONCLUSÃO	111
REFERÊNCIAS	113

1. INTRODUÇÃO

A descoberta de novas entidades químicas (*NCE*, do inglês *New Chemical Entities*) é um processo longo, interdisciplinar, complexo e que envolve altos investimentos (BRODNIEWICZ, GRYNKIEWICZ, 2010). As etapas envolvidas desde a proposta original de um fármaco até seu lançamento podem levar de 12 a 15 anos e ter um custo superior a US\$ 1 bilhão (HUGHES, 2011). De todas as etapas necessárias para introduzir um novo fármaco no mercado, os ensaios clínicos proporcionam a etapa mais cara, representando 63% do custo total (LI, AN, JONES, 2011). Visando reduzir custos e o tempo empregado nessas etapas, a indústria farmacêutica mais do que dobrou seu investimento em pesquisa e desenvolvimento desde 1998. Entretanto, o número de produtos lançados nesse mesmo período não seguiu o mesmo ritmo (Figura 1) (MEANWELL, 2011).

Figura 1. Relação entre os investimentos em Pesquisa e Desenvolvimento feitos pela indústria farmacêutica e fabricantes da América. Em função do número de novas entidades químicas (NCE) e imunobiológicos.



Fonte: Adaptado de MEANWELL, 2011.

Entre as estratégias adotadas para reduzir custos e tempo, principalmente nas etapas iniciais de desenvolvimento de fármacos, pode-se destacar a técnica de reposição (do inglês *repurposing*), que utiliza moléculas com atividade terapêutica estabelecida para novas indicações. A grande vantagem dessa estratégia é a redução do custo e tempo necessários para o desenvolvimento dos fármacos, uma

vez que suas propriedades farmacocinéticas, segurança e eficácia já são conhecidos. Contudo deve-se destacar que essa é uma solução temporária para o problema, uma vez que o arsenal terapêutico existente não é ampliado e que muitas empresas farmacêuticas não permitem um estudo amplo de seus fármacos, com receio da descoberta de efeitos colaterais desconhecidos (NWAKA, HUDSON, 2006, GUIDO, OLIVA, 2009, SAMS-DODD, 2006).

De forma análoga, pode-se utilizar moléculas bioativas identificadas nas etapas pré-clínicas de otimização como ponto de partida para o desenvolvimento de fármacos quando o alvo terapêutico presente no ser humano (indicação original) também está presente no agente causador da doença. Essa abordagem se diferencia da extensão da indicação terapêutica por exigir novos estudos de otimização dos protótipos frente ao seu novo alvo terapêutico (macromolécula ortóloga), enquanto na extensão da indicação, faz-se uso do fármaco propriamente dito para tratar outras doenças (GELB, 2007, GUIDO, OLIVA, 2009)

Por fim, é comum a combinação de fármacos já existentes para suprir necessidades terapêuticas ainda não satisfeitas (BURCHAN, 2011, LI, AN, JONES, 2011). Geralmente as combinações são empregadas com o intuito de possibilitar um efeito sinérgico, redução da toxicidade, redução do tempo de tratamento e diminuição do desenvolvimento de resistência (PINK et al, 2005).

Tendo em vista que somente a segunda estratégia permite, de fato, ampliar o arsenal terapêutico contra doenças tropicais negligenciadas e que ela não exclui as etapas pré-clínicas de otimização dos compostos protótipos, segue uma breve descrição das principais etapas e metodologias empregadas para avaliar e otimizar a potência de moléculas candidatas a fármacos, bem como alguns dados sobre a doença tropical negligenciada que será foco desse trabalho, a doença de Chagas.

1.1 PROCESSO DE IDENTIFICAÇÃO E OTIMIZAÇÃO DE MOLÉCULAS BIOATIVAS

O processo de desenvolvimento de fármacos pode ser dividido nas fases pré-clínica e clínica. As abordagens e técnicas empregadas na fase pré-clínica geralmente são guiadas por informações oriundas do alvo biomacromolecular e/ou na resposta fenotípica em um modelo da doença estudada, no qual se avalia a atividade de uma molécula potencialmente bioativa (LOMBARDINO, LOWE, 2004).

Ensaio de atividade biológica baseados em respostas fenotípicas buscam encontrar moléculas que proporcionem o efeito terapêutico, regulando uma anormalidade funcional específica da doença investigada ou reduzindo seus sintomas ou mudanças fisiológicas (SAMS-DODD, 2006). Para tal objetivo, os compostos são testados contra células, tecidos ou em um modelo animal e selecionados por sua capacidade de produzir o fenótipo de interesse no modelo em estudo (SWINNEY, ANTHONY, 2011, SAMS-DODD, 2006). Entretanto, o desconhecimento do mecanismo de ação pode levar a seleção de moléculas que tem baixa eficiência em seres humanos, pois o modelo de doença utilizado pode ter pouca significância clínica. Um outro problema relacionado a questão da incerteza do alvo terapêutico de ação da substância é a impossibilidade de direcionar um composto específico ao alvo, podendo conduzir aos testes substâncias com toxicidade previamente desconhecida a médio e longo prazo. Devido a esses fatores, a abordagem fenotípica foi sendo, paulatinamente restringida na descoberta fármacos, fazendo com que abordagens alternativas fossem desenvolvidas (BROWN, 2007).

Com o advento da tecnologia do DNA recombinante e subsequente sequenciamento do genoma de diversos organismos, novas proteínas foram sendo descobertas e estudadas. Como consequência, passou-se a utilizar com maior frequência informações estruturais das biomacromoléculas, juntamente com o conhecimento dos mecanismos bioquímicos associados a determinada doença para buscar alternativas terapêuticas que resultassem na prevenção e/ou cura. Essa abordagem parte do pressuposto que a modulação da atividade de uma biomacromolécula, essencial para o mecanismo patológico em questão, pode ocasionar o efeito terapêutico desejado. Assim, moléculas que se ligam ao alvo terapêutico, ativando-o ou inibindo-o, fornecem a base racional para a descoberta e otimização de protótipos a fármacos. Nessa perspectiva, a identificação e validação do alvo biomacromolecular constitui a etapa primordial do processo de planejamento de novos fármacos (SWINNEY, ANTHONY, 2011, SAMS-DODD, 2006).

A identificação do alvo terapêutico pode acontecer de maneira direta, através de métodos de purificação por afinidade, como cromatografia, fracionamento bioquímico e ligante radioativo ou indireta, através da utilização de banco de dados genômicos e perfil proteômico (TITOV, LIO, 2012; TASHIRO, IMOTO, 2012).

A etapa posterior a identificação do alvo terapêutico é a sua validação. A validação é importante pois se a inativação do alvo selecionado não produz a atividade biológica desejada, recursos desnecessários serão aplicados em estudos pré-clínicos como também na etapa dos ensaios clínicos (HOELDER, CLARKE, WORKMAN, 2012). A validação do alvo terapêutico pode ser realizada com modelos que não expressam a macromolécula em questão, que a expressam parcialmente ou através da utilização de um inibidor seletivo, caso este esteja disponível (SWINNEY, ANTHONY, 2011, SAMS-DODD, 2006).

Na sequência, diversas técnicas podem ser empregadas para selecionar e otimizar substâncias biologicamente ativas. A triagem biológica automatizada em alta escala (HTS, do inglês *high throughput screening*) e a triagem virtual (VS, do inglês *virtual screening*) são utilizadas para explorar o espaço químico centrado em determinados alvos biológicos (FERREIRA, OLIVA, ANDRICOPULO, 2011). As técnicas de HTS permitem testar cerca de 10 a 100 mil compostos por dia contra proteínas alvo enquanto que o ultra HTS avalia mais de 100 mil compostos (MAYR, BOJANIC, 2009). Já o VS é um método *in silico* empregado para selecionar substâncias potencialmente ativas, oriundas de um banco de dados, para posterior avaliação em testes *in vitro* (FERREIRA, OLIVA, ANDRICOPULO, 2011).

Nos estágios pré-clínicos é priorizada a otimização dos compostos líderes em termos de propriedades como potência, seletividade, absorção, distribuição, metabolismo e excreção, visando a obtenção de moléculas com propriedades farmacodinâmicas e farmacocinéticas adequadas para prosseguir na etapa clínica (GUIDO, ANDRICOPULO, OLIVA, 2010, SHEN, 2010). A fim de auxiliar no processo de otimização de compostos líderes, pode-se fazer o uso de ferramentas computacionais que auxiliam a estabelecer a relação entre a estrutura química e a atividade biológica de compostos (SAR, do inglês *structure-activity relationship*) .

1.2 FERRAMENTAS COMPUTACIONAIS NO DESENVOLVIMENTO DE FÁRMACOS

O desenvolvimento de moléculas biologicamente ativas, com propriedades farmacocinéticas e farmacodinâmicas otimizadas, é uma estratégia multidisciplinar que envolve a integração de uma grande variedade de tecnologias (GUIDO et al, 2008). O uso de técnicas computacionais no processo de descoberta e

desenvolvimento de fármacos está rapidamente ganhando popularidade e apreciação, pois permite reduzir custos e agilizar as etapas pré-clínicas de desenvolvimento de fármacos. Esse objetivo é alcançado através da integração de informações sobre o espaço químico e biológico das substâncias em estudo (SCHEIDER, 2011; D'OCA, 2011). Por isso, métodos de quimio- e bioinformática são ferramentas amplamente empregadas desde a fase de identificação da molécula bioativas até a otimização das propriedades físicoquímicas que levam ao aumento da potência e seletividade (JÓNSDÓTTIR, FLEMMING, BRUNAK, 2005).

No entanto, determinar qual ferramenta computacional utilizar durante o planejamento de fármacos depende, em grande parte, dos dados disponíveis sobre o alvo macromolecular e/ou ligantes bioativos (ANDRICOPULO et al, 2009). Quando a estrutura tridimensional da biomacromolécula não está disponível, a estratégia para a descoberta, desenvolvimento e otimização de fármacos pode ser baseada em ligantes conhecidos da proteína estudada (LBDD, do inglês *ligand-based drug design*). Abordagens baseadas em modelos farmacofóricos, busca por similaridade molecular e o estudo das relações quantitativas entre a estrutura química e a atividade biológica (QSAR, do inglês *quantitative structure-activity relationships*) são exemplos de metodologias baseada no ligante (LEE, HUANG, JUAN, 2011, VIRTANEN, PENTIKAINEN, 2010).

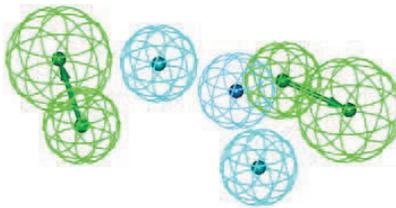
Farmacóforo é um conjunto de características estereais e eletrônicas comuns a um conjunto de moléculas ativas, necessário para a interação com o alvo biológico e exigido para proporcionar a atividade biológica. Essa abordagem tem sido amplamente utilizada tanto no processo de descoberta de compostos protótipos, através do VS, quanto para otimização. Na abordagem VS, o modelo farmacofórico auxilia na identificação de substâncias, em banco de dados, que apresentam as mesmas características essenciais para a atividade encontrada nas moléculas ativas utilizadas como referência (DRIE, 2007, YANG, 2010).

A busca por similaridade molecular é uma metodologia baseada na premissa de que moléculas com similaridade química frequentemente exibem atividade biológica similar. Ela difere dos modelos farmacofóricos, pois nessa metodologia todos os átomos ou moléculas são considerados para o cálculo da similaridade, enquanto no modelo farmacofórico somente os grupos farmacofóricos são considerados (Figura 2) (JORGENSEN, 2010).

Figura 2. Exemplo de processos baseados no ligante. No modelo farmacofórico, um conjunto de características estéreas e eletrônicas em comum de um conjunto de dados são selecionadas por serem importantes na interação com o alvo macromolecular. No modelo por similaridade química, representado pela similaridade morfológica do programa surflex-sim, cria-se pontos de observação na molécula referência de acordo com as características de Van der Waals e ligações de hidrogênio. Um conjunto de pontos externos (triângulo) são formados, com a área representando as características moleculares e comparados com outros triângulos em busca de características moleculares semelhantes.

Processos de descoberta de fármacos baseado no ligante 3D

Modelo farmacofórico

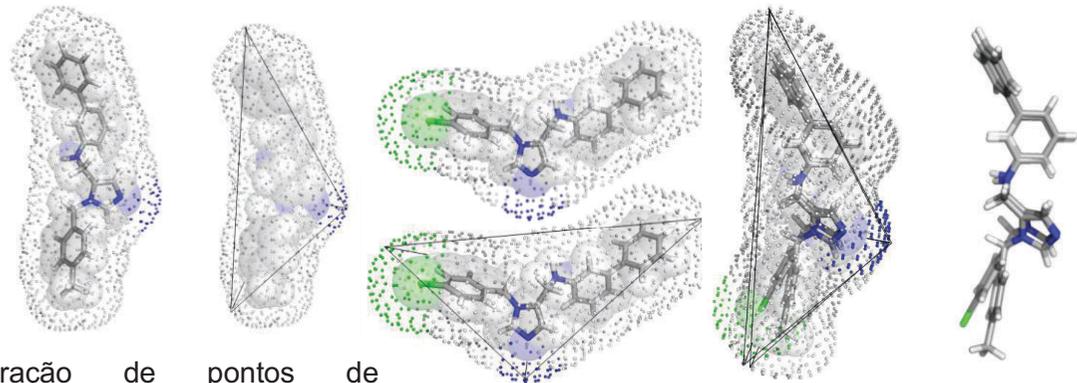


Representação de características importantes em um grupo de dados. Verde representa aceptores de hidrogênio; azul claro características aromáticas; ciano para características hidrofóbicas.



Moléculas no modelo farmacofórico.

Modelo de similaridade química



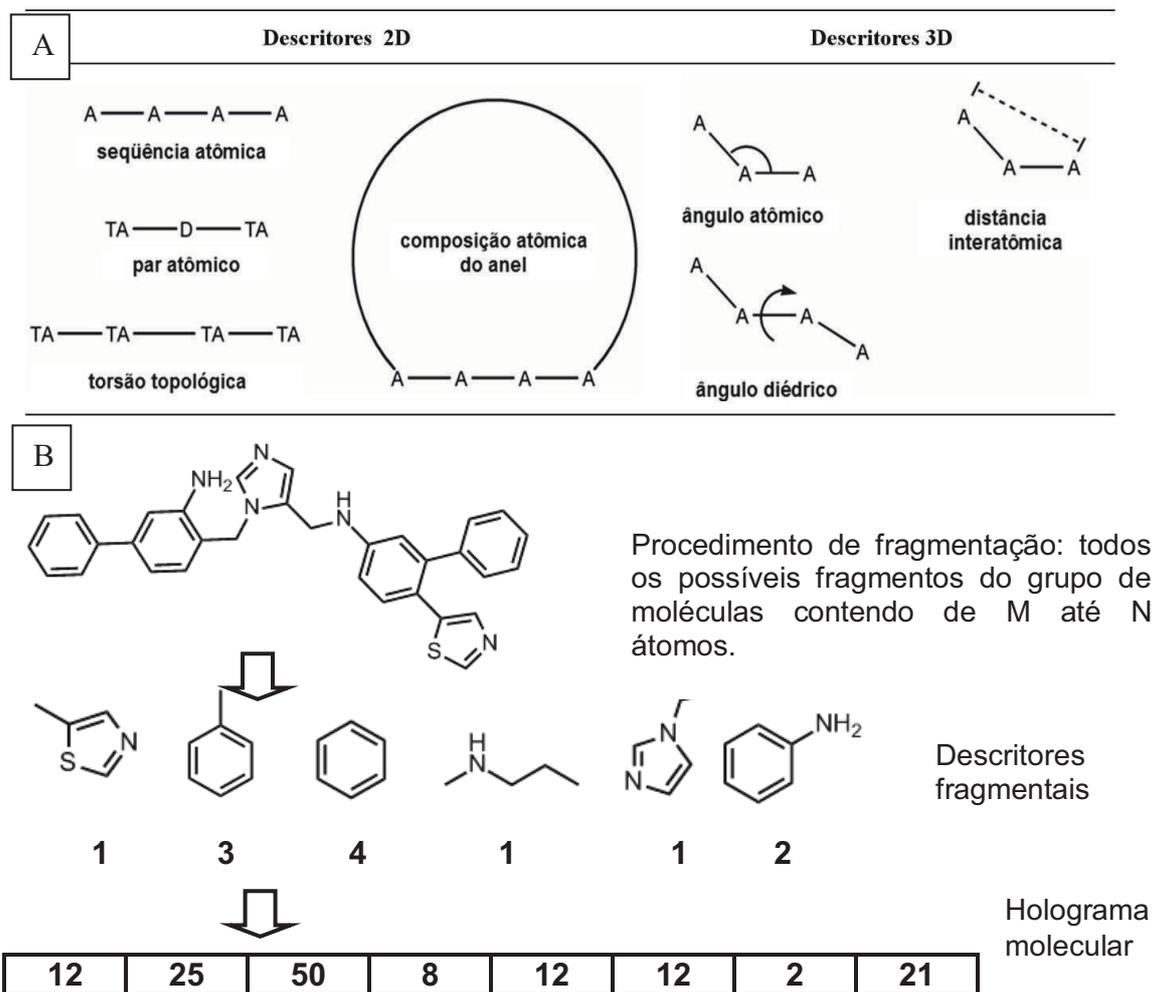
Geração de pontos de observação na molécula referência de acordo com as características da molécula como um todo.

Molécula de entrada

Comparação dos pontos de observação baseado na molécula inicial e alinhamento através da semelhança

O QSAR permite uma investigação ampla das propriedades estruturais e físico-químicas que determinam a atividade biológica de uma substância. Essas propriedades são codificadas em descritores, que podem ser fragmentais, físico-químicos, topológicos, constitucionais, etc. Os índices topológicos, por exemplo, permitem representar o tamanho, ramificação e a forma global da estrutura molecular, desprezando a natureza dos átomos; índices constitucionais, por sua vez, revelam a natureza dos átomos envolvidos e os índices de propriedade proporcionam a distinção das propriedades físico-químicas da molécula estudada (ACI, 2011). Assim, a utilização de diferentes tipos de descritores permite descrever de forma apropriada as características moleculares que se relacionam com a atividade biológica (Figura 3).

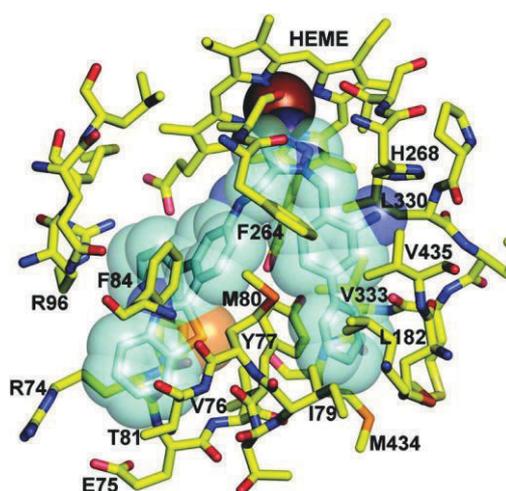
Figura 3. Descritores moleculares utilizados no desenvolvimento de modelos de QSAR. Representado em A descritores topológicos, constitucionais, físico-químicos etc, usados em modelos como QSAR baseado em descritores bidimensionais. Em B, descritores fragmentais, usados, por exemplo, na metodologia de HQSAR com a utilização de hologramas.



Adaptado WANG, 2003, WANG 2009, DUDEK, ARODZ, GÁLVEZC, 2006.

Com o conhecimento das estruturas de alvos macromoleculares, é possível a aplicação de estratégias de planejamento de fármacos baseado na estrutura do receptor (SBDD, do inglês, *structure-based drug design*). O avanço da genômica e proteômica, aliados à evolução das técnicas de cristalografia de raios X e ressonância magnética nuclear (RMN) proporcionam um aumento significativo no número de alvos moleculares com estruturas 3D disponíveis no Banco de Dados de Proteínas. (GUIDO, ANDRICOPULO, OLIVA, 2010). Nessa situação, pode-se aplicar técnicas computacionais, como o acoplamento molecular (Figura 4), para analisar as propriedades estéreas e eletrônicas dos bolsões do alvo e assim definir interações entre ligante-receptor que sejam relevantes para a propriedade biológica dos compostos.

Figura 4. Representação da metodologia do acoplamento molecular. Interações entre o ligante e os aminoácidos presente no sítio ativo da Lanosterol 14 alfa-desmetilase são analisados.

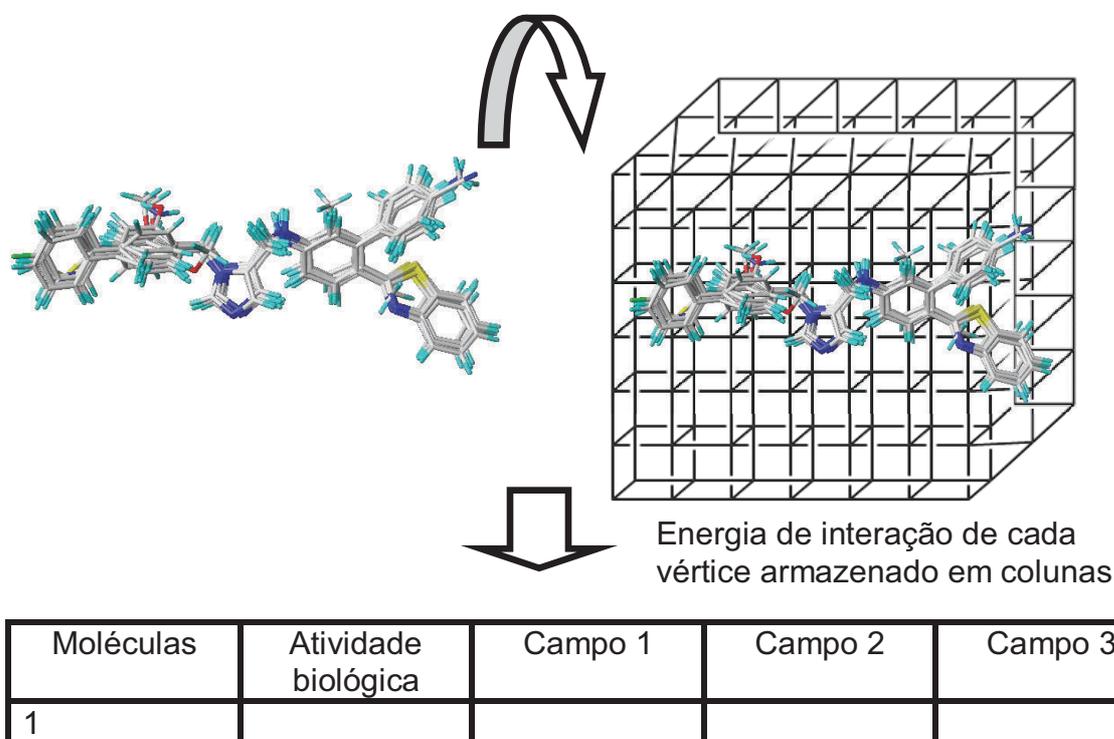


Fonte: SURYADEVARA, 2009.

Em muitos casos, o uso integrado de estratégias de SBDD e LBDD pode gerar informações complementares para o planejamento de novas moléculas bioativas (GUIDO, ANDRICOPULO, OLIVA, 2010). A Análise Comparativa de Campos Moleculares (CoMFA, do inglês *Comparative Molecular Field Analysis*) é um exemplo claro de técnica LBDD que pode se beneficiar de informações estruturais acerca do alvo molecular. O CoMFA é um método de QSAR 3D que utiliza campos de energia estérea (Van der Waals) e eletrostática (Coulomb) para descrever interações favoráveis e desfavoráveis nos compostos de interesse (CRAMER, DAVID, BUNCE, 1988). Neste método, as moléculas em estudo são alinhadas em

uma grade tridimensional e as interações eletrostáticas e estéricas de uma carga de prova com cada uma das moléculas são medidas. Embora os campos estéricos e eletrostáticos representem a natureza entálpica da interação entre fármaco e proteína, a componente entrópica, como a dessolvatação ou a diminuição da liberdade de rotação, é descartada durante a análise (AKAMATSU, 2002, GOHDAA, 2000). Os valores de energia de interação (variáveis independentes) podem ser correlacionados com a afinidade de ligação ou outros valores de atividade biológica (variável dependente) através de regressão por mínimos quadrados parciais (PLS) (PETERSON, SCHAAL, KARLÉN, 2004) (Figura 5).

Figura 5. Etapas envolvidas na geração do modelo de CoMFA.



Fonte: Adaptação CRAMER, PATTERSON, BUNCE, 1988.

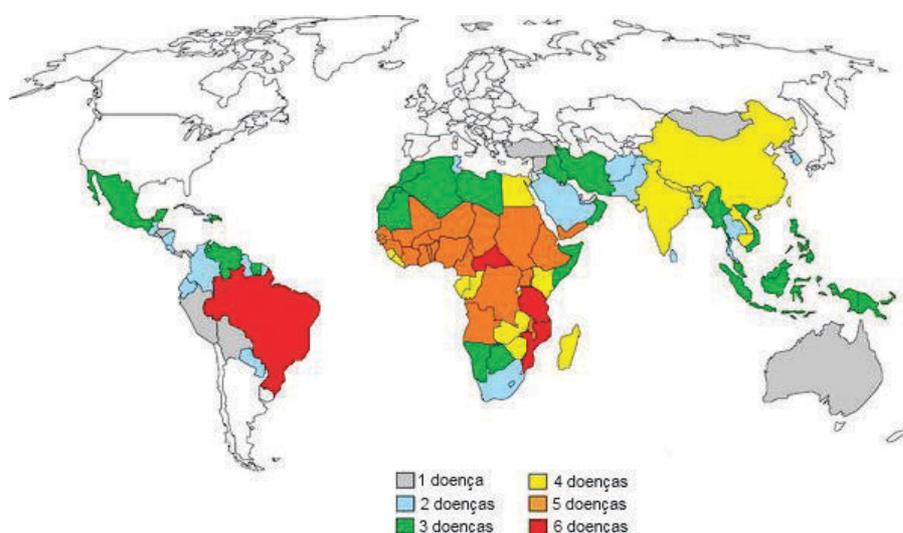
As técnicas computacionais descritas acima podem ser úteis para acelerar as etapas pré-clínicas de planejamento de fármacos, o que é imprescindível para projetos voltados para investigação de alternativas terapêuticas contra doenças negligenciadas para as quais ainda não existe tratamento eficiente (CHATELAIN, IOSET, 2011, PINK et al, 2005). A fim de compreender a problemática envolvida no planejamento de fármacos contra doenças que afetam milhões de pessoas em países subdesenvolvidos ou em desenvolvimento, segue um breve panorama sobre o impacto socioeconômico dessas doenças, com enfoque especial na doença de Chagas.

1.3 DOENÇAS TROPICAIS NEGLIGENCIADAS

As doenças negligenciadas estão associadas, principalmente, a infecções bacterianas, parasitárias, fúngicas ou virais que assolam as camadas mais pobres da população de países em desenvolvimento (HOTEZ et al, 2008), particularmente aqueles localizados nas regiões de clima tropical, razão pela qual elas também podem ser denominadas de doenças tropicais negligenciadas (BEYRER, 2007).

Por ser predominantemente de países subdesenvolvidos e em desenvolvimento, o controle e eliminação dessas doenças tornaram-se uma das metas da Organização Mundial da Saúde (OMS) para a diminuição da pobreza mundial. Para que isso aconteça, a OMS recomenda cinco estratégias de saúde pública para a prevenção e controle das doenças negligenciadas: a) controle do vetor, b) controle no fornecimento de água potável, c) saneamento, higiene e saúde pública veterinária, d) intensificação da gestão em determinados casos, e) quimioterapia preventiva (OMS, 2010). Embora a profilaxia e terapia ocasionam um grande impacto na prevalência dessas doenças, elas continuam sendo endêmicas em muitos países (Figura 6) (MOLYNEUX, HOTEZ, FENWICK, 2006). Como consequência, doenças como malária, tuberculose, doenças diarreicas e kinetoplastideos (como a doença de Chagas), causam aproximadamente 35.000 mortes por dia nos países em desenvolvimento (COHEN, DIBNER, WILSON, 2010).

Figura 6. Representação do número de doenças negligenciadas por região do globo terrestre.



Fonte: Adaptado de MOLYNEUX, HOTEZ, FENWICK, 2006.

Diversos fatores limitam a utilidade dos medicamentos existentes, como o alto custo, baixa taxa de adesão ao tratamento por parte dos pacientes, resistência dos agentes infecciosos aos fármacos existentes e baixa eficácia dos medicamentos disponíveis. Diante disso, fica evidente a importância de novas formas de diagnóstico e de agentes terapêuticos mais eficientes e menos tóxicos, que auxiliem na prevenção e tratamento de pacientes infectados ou sob risco de infecção (PINK et al, 2005, NWAKA, HUDSON, 2006). Entretanto, os orçamentos limitados na maioria das nações em desenvolvimento, o baixo poder de compra do público em geral e a baixa probabilidade de retorno financeiro satisfatório fez com que houvesse pouco interesse da indústria farmacêutica no desenvolvimento de fármacos e conjunto diagnóstico para essas doenças (COHEN, DIBNER, WILSON, 2010, PINK et al, 2005, NWAKA, HUDSON, 2006).

Embora doenças negligenciadas sejam responsáveis por 11,4% da carga global de doenças existentes no período entre 1975 e 2004, apenas 21 (1,3%) dos 1.556 novos medicamentos registrados foram desenvolvidos para essas doenças. (CHIRAC, TORREELE, 2006, COHEN, DIBNER, WILSON, 2010). Portanto, é clara a necessidade de um maior investimento no desenvolvimento de fármacos e conjuntos diagnósticos para doenças negligenciadas. A partir de 2000 houve um interesse renovado, pelas indústrias farmacêuticas, no desenvolvimento de fármacos para doenças negligenciadas (PINK et al, 2005; NWAKA, HUDSON, 2006) (Tabela 1). Isso se deve em grande parte ao estabelecimento de parcerias público-privadas (PPP) como a Medicines for Malaria Venture (MMV), Drugs for Neglected Diseases initiative (DNDi) e o Institute for One World Health (IOWH) (NWAKA, RIDLEY, 2003) que, contando com o apoio de empresas como a Pfizer, Merck Serono, Sanofi-Aventis, Bayer entre outras, financiaram direta ou indiretamente projetos para o desenvolvimento de fármacos contra doenças negligenciadas (PINK et al, 2005, NWAKA, HUDSON, 2006). O papel dessas companhias no processo de descoberta de novos fármacos se resume em realizar ensaios de triagens de alto desempenho e desenvolver modelos farmacocinéticos (GUIDO, OLIVA, 2009). Resumindo, a maioria das PPPs se dedica a otimização de compostos protótipos previamente identificados.

Tabela 1. Fármacos e imunobiológicos aprovados para doenças negligenciadas no período de 2000 a 2009.

Doenças	Fármacos ou vacinas	Ano de aprovação	Países de aprovação	
<i>HIV/AIDS</i>	Lamivudina + Zidovudina	2002	EUA	
	Lamivudina 3TC	2002	EUA	
	Estavudina D4T	2002	EUA	
	Ritonavir RTV	2005	EUA	
	Nelfinavir NPV	2004	EUA	
	Enfuvitida	2006	EUA	
	Efavirenz + Emtricitabine + Tenofovir	2006	EUA	
	Emtricitabina FTC	2003	EUA	
	Lopinavir + Ritonavir	2000	EUA	
	Fasamprenavir	2003	EUA	
	Tipranavir			
<i>Malária</i>	Atovaquona + Proguanil	2000	EUA	
	Artemisina + Amodiaquina	2008	Kenia	
	Artemotil	2000	Holanda	
	Artesunato	2001	Kenia	
	Diidroartemisinina	2001	Kenia	
	Diidroartemisinina + Piperaquine	2005	Kenia	
	Artesunato + Mefloquina	2005	Kenia	
	Artesunato + Amodiaquina	2007	Marrocos	
	Artesunato + Sulfadoxina + Pirimetamina	2007	Kenia	
	Diidroartemisinina + Piperaquina + Trimetoprima	2006	Kenia	
	Clorproguanil + Dapsona	2003	Kenia	
	<i>Tuberculose</i>	-----		
	<i>Pneumonia</i>			
	<i>Bacteriana e</i>			
<i>Meningite</i>	Menactra (vacina)	2005	EUA	
<i>Doenças</i>				
<i>diarreicas</i>	Vacina pentavalente para rotavírus	2006	EUA	

Tabela 1 (continuação). Fármacos e imunobiológicos aprovados para doenças negligenciadas no período de 2000 a 2009.

Doenças	Fármacos ou vacinas	Ano de aprovação	Países de aprovação
<i>Doenças diarreicas</i>	Dukoral (vacina para cólera e diarreia)	2004	Inglaterra
	Nitazoxanida	2004	EUA
<i>Kinetoplastida</i>			
<i>Leishmaniose</i>	Miltefosina	2002	Índia
<i>Úlcera de Buruli</i>	-----		
<i>Dengue</i>	-----		
<i>Helmintos</i>	-----		
<i>Lepra</i>	-----		
<i>Tracoma</i>	-----		
<i>Febre Reumática</i>	-----		
<i>Febre Tifóide e paratifóide</i>	-----		

Fonte: COHEN, DIBNER, WILSON, 2010.

Entretanto, fármacos para o tratamento da AIDS e malária representam 84,61% fármacos lançados no período de 2000 a 2009 respectivamente, enquanto que doenças causadas pelo grupo dos Kinetoplastidas, como a Doença de Chagas, apenas um fármaco foi desenvolvido (COHEN, DIBNER, WILSON, 2010). Portanto, o estudo de protótipos de fármacos para essa doença negligenciada se torna prioritário.

1.3.1 DOENÇA DE CHAGAS

A doença de Chagas foi reportada em 1909 pelo pesquisador Carlos Chagas, que descobriu a existência do protozoário *Trypanosoma cruzi* (*T. cruzi*), revelou o papel dos insetos triatomíneos no ciclo de infecção como hospedeiro intermediário, identificou quais animais são reservatórios silvestres de *T. cruzi* e comprovou ser esse protozoário o agente causador dessa doença (CHAGAS, 1909).

O ciclo de vida do parasito apresenta vários estágios de desenvolvimento tanto no inseto vetor triatomíneo quanto no hospedeiro humano. Tripomastigotas sanguíneos e amastigotas intracelulares são as formas típicas encontradas no

organismo do hospedeiro, enquanto que epimastigotas replicativas e tripomastigotas metacíclicas infectam os vetores triatomíneo (RASSI, RASSI, MARIN-NETO, 2010). Normalmente, a infecção pelo protozoário *T. cruzi* ocorre via vetorial, quando o inseto excreta o parasita durante o repasto sanguíneo, penetrando a mucosa da pele lesionada. Entretanto, a doença pode ser transmitida por mecanismo não vetorial, como por via congênita, transfusão sanguínea, transplante de órgãos, via oral ou em acidente de laboratório (EPTING, COATES, ENGMAN, 2010).

A penetração da forma infectante tripomastigota pode levar uma reação denominada chagoma de inoculação, com edema local e endurecimento, congestão vascular e infiltração celular (RASSI, RASSI, MARIN-NETO, 2010). Depois da infecção inicial pela forma de tripomastigota, os pacientes desenvolvem a fase aguda da doença, que usualmente é assintomática. Quando sintomático, o indivíduo pode apresentar febre, aumento do tamanho do fígado e do baço, insuficiência cardíaca devido à inflamação do miocárdio (miocardite), efusão do pericárdio, hepatoesplenomegalia, linfadenopatia e, menos frequentemente, meningoencefalite. Manifestações da fase aguda resolvem-se espontaneamente em cerca de 90% dos indivíduos infectados (RASSI, RASSI, MARIN-NETO, 2010; EPTING, COATES, ENGMAN, 2010). Cerca de 60 a 70% desses indivíduos nunca desenvolvem a doença clinicamente aparente. Os 30-40% restantes desenvolvem a forma crônica da doença, seja na forma cardíaca, digestiva (megaesôfago e megacólon), cardiodigestiva (MAYA et al, 2010) ou neurológica (OLIVEIRA, 2011). O impacto social e econômico da fase crônica da doença de Chagas pode ser calculado através do número de anos de vida perdidos em função da incapacidade decorrente da doença (do inglês *DALY – disability adjusted life years*) que, para pacientes chagásicos é de aproximadamente 700 mil (HOTEZ et al, 2006, WHO, 2010).

O controle da transmissão vetorial da doença de Chagas teve início com a melhoria da qualidade das construções de casas domiciliares e pulverização de inseticidas nas paredes das casas em doses relativamente altas (VILLELA et al, 2009, DIAS, SILVEIRA, SCHOFIELD, 2002). Como resultado desse tipo de estratégia de controle da doença de Chagas, o Brasil recebeu em 2006 a certificação internacional de interrupção da transmissão da doença pelo *T. infestans*, concedida pela Organização Panamericana da Saúde e Organização Mundial da Saúde. (YAMAGATA, NAKAGAWA, 2006, MASSAD, 2008). Entretanto, uma das grandes dificuldades para o controle da doença de Chagas é a existência de outras

espécies de vetores silvestres que podem invadir e reinfestar o nicho que estava sendo ocupado pelo *T. infestans* (Figura 7) (COURA, 2007, YAMAGATA, NAKAGAWA, 2006, ABDAD-FRANCHA, 2009).

Figura 7. Diversidade de espécies de insetos triatomíneos (linhagem Rhodniini) por unidade taxonômica operacional (OTU, do inglês *operational taxonomic units*) na América Latina.



Fonte: Adaptado por ABDAD-FRANCHA, 2009.

Hoje, entre as diversas doenças negligenciadas causadas por protozoários encontradas na América Latina, a doença de Chagas é a que apresenta maior impacto, afetando mais pessoas nessa região do que qualquer outra doença parasitária, incluindo a malária (DNDI, 2010; MORAN et al, 2009). A Doença de Chagas é uma patologia endêmica em 21 países da América Latina, provoca aproximadamente 8 milhões de casos e 14 mil mortes (DNDI, 2010). Contudo, o parasito *T. cruzi* tem sido encontrado também em países não endêmicos da América do Norte (Estados Unidos, onde cerca de 300 000 pessoas infectadas e Canadá, com aproximadamente 5.500), região do pacífico Oeste (particularmente Japão, com aproximadamente 3.000 casos e Austrália, com 1.500 casos) e mais recentemente na Europa (principalmente Bélgica, Espanha, França, Itália, Reino Unido e Suíça e em menor extensão Alemanha, Áustria, Croácia, Dinamarca, Holanda, Luxemburgo, Noruega, Portugal, Romênia e Suécia, com aproximadamente 80.000 casos (Figura 8) (COURA, VIÑAS, 2010, SHIKANAI-YASUDA, CARVALHO, 2012). O número de casos entre pessoas desses países que viajam para a América Latina tem

aumentado e para agravar a situação muitos pacientes desconhecem que adquiriram a doença e como consequência disso contribuem para sua disseminação em áreas não-endêmicas (COURA, VIÑAS, 2010, DNDI, 2010).

Figura 8. Rota de migração da América Latina e estimação do número total de indivíduos infectados em países não endêmicos.



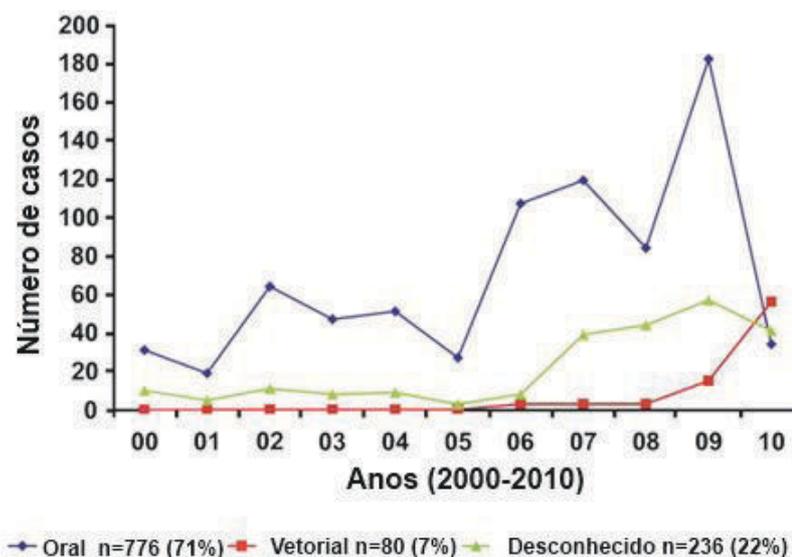
Fonte: COURA, VIÑAS, 2010.

No Brasil, a estimativa de pessoas infectadas é de aproximadamente 2 a 3 milhões, com incidência de 100 novos casos por ano (PONTES, et al, 2010). O perfil epidemiológico brasileiro da doença apresenta ocorrências de casos e surtos na Amazônia Legal por transmissão oral, devido a ingestão de alimentos contaminados (caldo de cana, açaí, bacaba, entre outros) e vetorial (sem colonização e extradomiciliar). No período de 2000 a 2010, foram registrados no Brasil 1.086 novos casos de doença de Chagas aguda. Destes, 71% (765/1.086) foram por transmissão oral, 7% por transmissão vetorial (80/1.086), em 22% (234/1.086) não foi identificada a forma de transmissão (SHIKANAI-YASUDA, CARVALHO, 2012) (Figura 9). Alguns surtos por via oral foram relatados na literatura como:

1. 11 casos em Belém do Pará no ano de 2000 (PINTO et al, 2009);
2. 11 casos no Pará em 2006 (NÓBREGA et al, 2009);
3. 7 casos em Macaúbas e 6 casos em Ibipitanga, ambas localizadas na região central do estado da Bahia, no período de 2006 (DIAS et al, 2008, BASTOS et al, 2010);
4. 9 casos em Tefé no ano de 2004 e 25 casos em Coari em 2007, cidades da região amazônica (BARBOSA-FERREIRA et al, 2010);
5. 24 casos em Santa Catarina no período de 2005 (ANDRADE et al, 2011);

6. 12 casos em Breves e 13 em Bagre, ilhotas do estado do Pará no período de 2007 (BELTRÃO et al, 2009).

Figura 9. Número de casos de pessoas infectadas pelo *Trypanosoma cruzi* nos anos de 2000 a 2010 no Brasil. Observa-se que, dentre os tipos de contaminação, a transmissão por via oral é a predominante.

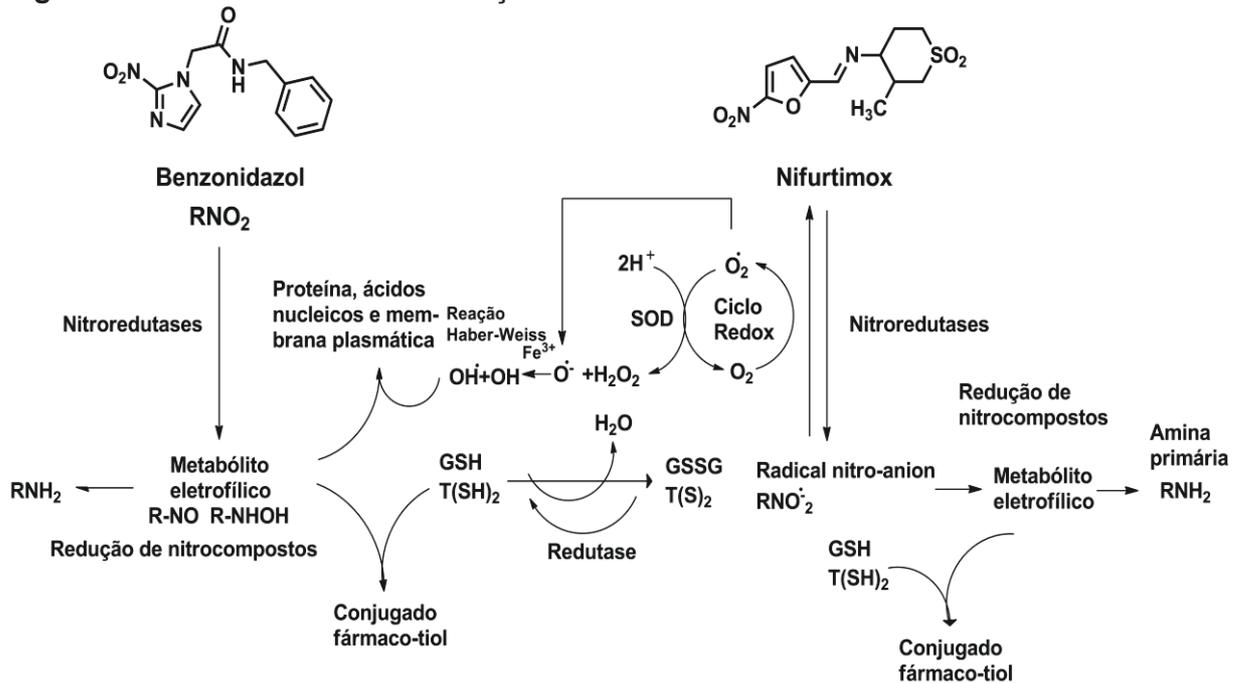


Fonte: Adaptado de SHIKANAI-YASUDA, CARVALHO, 2012.

Apesar do grande impacto que a doença de Chagas tem nos indicadores de saúde da América Latina, o arsenal terapêutico para seu tratamento é bastante limitado. Os fármacos disponíveis para o tratamento da doença de Chagas, nifurtimox e benzonidazol, atuam apenas na fase aguda da doença (MAYA et al, 2007; PEREZ-MOLINA, 2009) sendo que já existem relatos de resistência dos dois fármacos (MEJIA et al, 2010, WILKINSON et al, 2008). A ação tripanomicida de nifurtimox e benzonidazol se deve a formação de radicais nitro e radicais hidroxilas, respectivamente, que se ligam covalentemente a proteínas, ácidos nucleicos e membranas plasmáticas do parasita, levando-o a morte (Figura 10). No entanto, o mecanismo de ação antiparasitário desses fármacos também é responsável pelos efeitos colaterais (náuseas, vômitos, alterações hematológicas por hipersensibilidade (leucopenia, trombocitopenia) e até polineuropatia dose dependente) observados em pacientes que utilizam essa terapia (APT, ZULANTAY, 2011).

Apesar dos dados relatados acima, o investimento da indústria farmacêutica na descoberta e desenvolvimento de novos fármacos contra a doença de Chagas foi, durante muitos anos, praticamente inexistente.

Figura 10. Provável mecanismo de ação dos fármacos benzonidazol e nifurtimox.



Fonte: Adaptado de MAYA et al, 2007.

1.3.1.1 DESENVOLVIMENTO DE FÁRMACOS CONTRA DOENÇAS DE CHAGAS

A fase crônica da doença de Chagas tem grande importância epidemiológica na América Latina e em alguns estados desenvolvidos. No Brasil, embora a transmissão pelo vetor e por transfusão de sangue tenha diminuído, um grande número de indivíduos permanecem infectados (HASSLOCHER-MORENO, 2012, SCHMUNIS, YADON, 2010). Esse cenário, juntamente com o aparecimento da resistência dos fármacos atuais, aliado a sua baixa eficácia na fase crônica e altos efeitos colaterais, o desenvolvimento de novos fármacos antichagásico torna-se essencial (SILVA, et al, 2012, MEJIA, et al, 2012). Apesar disso, o desenvolvimento de novas abordagens para o tratamento da fase crônica da doença de Chagas durante muito tempo foi considerado irrelevante devido a teoria de que a sintomatologia seria ocasionada por reações autoimune (URBINA, 2010; URBINA, DOCAMPO, 2003; CROFT et al, 2005). Esta hipótese foi baseada na ausência aparente de parasitas nas lesões do coração e trato gastrointestinal, associadas à fase crônica sintomática da doença de Chagas e a presença de respostas a auto anticorpos em pacientes com a doença (CUNHA-NETO, KALIL, 1995). De acordo com essas observações, o tratamento com quimioterápicos não poderia conduzir a uma melhoria na evolução clínica dos pacientes portadores da sintomatologia

chagásica crônica. Entretanto, a hipótese auto-imune tem sido fortemente contestada, pois:

1. Foi observada a presença de parasitas nas lesões inflamatórias em pacientes chagásicos com comprometimento do trato gastrointestinal ou cardíaco (VAGO, 2000, SCHIJMAN et al, 2004, BENVENUTI et al, 2008)
2. Existe correlação entre a carga do antígeno parasitário, a intensidade dos processos inflamatórios e a gravidade da doença em modelos experimentais (ZHANG, TARLETON, 1999, PEREZ-FUENTES et al, 2003, BENVENUTI et al, 2008).
3. Há baixa correlação entre os níveis de auto-anticorpos e a sintomatologia em modelos experimentais na fase crônica da doença de Chagas (ZHANG, TARLETON, 1999, MARIN-NETO et al, 2008).
4. Ocorre exacerbação da doença em modelos experimentais e humanos imunocomprometidos (VAIDIAN et al, 2004, FIORELLI et al, 2005, DIEZ et al, 2007, CORDOVA et al, 2008).
5. Observa-se efeito positivo da quimioterapia específica sobre a evolução da fase crônica da doença (GARG, TARLETON, 2002, TARLETON, 2003, CORRALES et al, 2005, GARCIA et al, 2005).

Como a quimioterapia existente para a doença não é efetiva na fase crônica, torna-se necessário identificar novos alvos macromoleculares para o desenvolvimento de alternativas terapêuticas mais eficientes. Exemplos de rotas metabólicas e enzimas que estão sendo exploradas nesse sentido são:

1. Inibidores da via de recuperação de purinas – Os tripanosomatídeos são auxotróficos para purinas, ou seja, eles não realizam sua síntese *de novo*. Dependendo, portanto, de purinas pré-formadas do hospedeiro para sua sobrevivência. Uma enzima chave nesse processo é a hipoxantina-guanina fosforibosil transferase (HGPRT), que promove a incorporação de alopurinol ao DNA do parasita, interrompendo assim sua síntese protéica e de RNA (APT, 2010).
2. Inibidores da síntese e metabolismo de tripanotiona – essa via bioquímica, responsável pela proteção do parasita contra o estresse oxidativo, é exclusiva

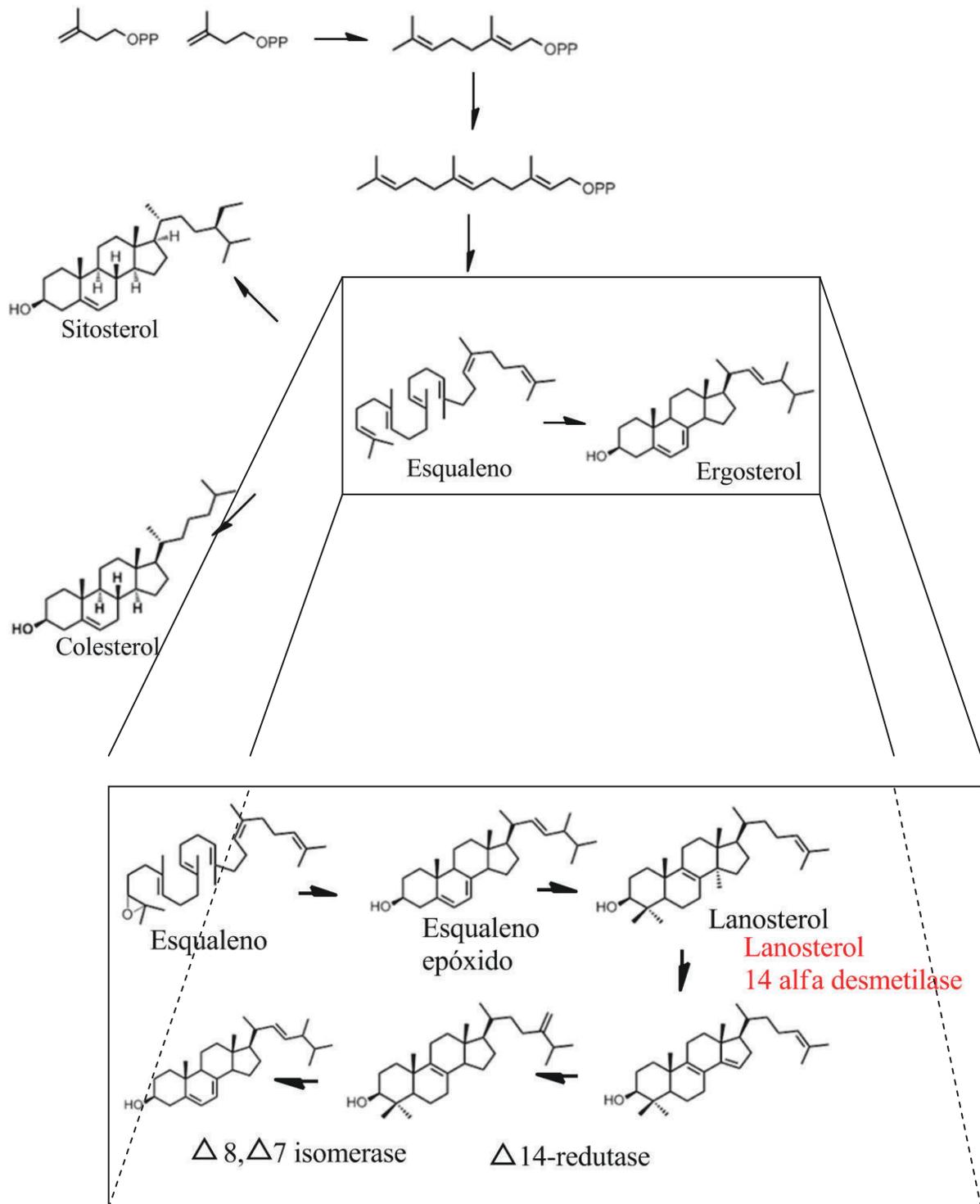
A rota biossintética desses esteróis inicia-se com unidades de acetato e envolve a enzima hidroximetilglutaril CO-A. Posteriormente, é convertido em ácido mevalômico, farnesil pirofosfato e esqualeno epoxidase. A partir do esqualeno, as rotas biossintéticas divergem entre os reinos, produzindo esteróides específicos: nos animais o principal esteróide sintetizado é o colesterol, enquanto nos fungos encontra-se principalmente o ergosterol e em plantas superiores o sitosterol (GUBBINS, ANAÏSSIE, 2009, MOURITSENA, ZUCKERMANN, 2004).

A partir da descoberta que *T. cruzi* apresenta ergosterol na sua membrana e não colesterol, enzimas importantes dessa via biossintética foram reconhecidas como alvos terapêuticos potenciais para o desenvolvimento de fármacos antichagásicos (BEACH, GOAD, HOLTZ, 1986). Um importante passo para a biossíntese do ergosterol envolve a remoção do grupo metil do composto 14-metil lanosterol, intermediário da rota biossintética do ergosterol, pela enzima lanosterol 14 alfa-desmetilase, pertencente a superfamília do citocromo P₄₅₀ (Figura 11) (SURYADEVARA et al, 2009, LEPESHEVA, 2006).

A enzima lanosterol 14 alfa-desmetilase de *T. cruzi* é uma heme proteína caracterizada por um conjunto de 13 hélices α e 12 fitas β . O seu sítio ativo é bastante hidrofóbico, composto pelas hélices B'/alça B'C, hélice C, hélice I e uma alça (CHEN et al, 2010). Além do sítio ativo, existe um caminho de resíduos hidrofóbicos que permite o acesso do ligante ao sítio ativo. Esse túnel hidrofóbico é caracterizado por conter aminoácidos presentes na hélice H, hélice F" e uma alça (Figura 12) (LEPESHEVA et al, 2010).

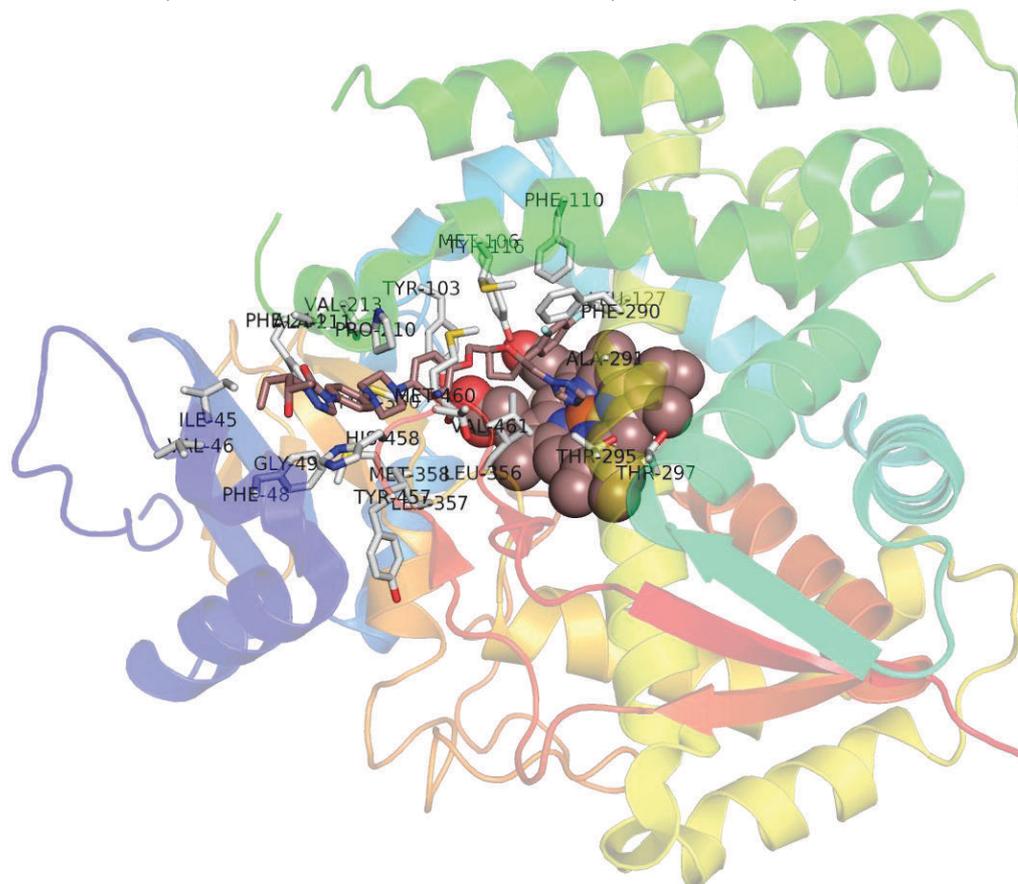
Devido ao fato dos esteróis serem essenciais para a viabilidade e proliferação de determinadas espécies, a inibição dessa via biossintética foi bastante explorada para a descoberta de fármacos (ISHIDA et al, 2011). Como exemplo, pode-se citar os antifúngicos usados na clínica que inibem enzimas da via de síntese do ergosterol (ODDS, BROWN, 2003).

Figura 11. Via biossintética resumida do ergosterol.



Fonte: Adaptação GUBBINS, ANAISSIE, 2009, MOURITSENA, ZUCKERMANN, 2004, LEPESHEVA, 2007.

Figura 12. Sítio ativo e bolsão hidrofóbico da enzima lanosterol 14 alfa-desmetilase de *T. cruzi* com o inibidor posaconazol. O sitio ativo está representado pela hélice B'/alça B'C, de coloração azul claro, onde situa-se os aminoácidos tirosina 103, metionina 106, fenilalanina 110 e tirosina 116; pela hélice C, de coloração azul claro e situada atrás da figura (não visível), que contém o aminoácido leucina 127; hélice I, de coloração amarela, onde estão presentes os aminoácidos alanina 297, fenilalanina 290, alanina 291 e tirosina 295 e uma alça representada em laranja, com a presença dos aminoácidos leucina 356, leucina 357 e metionina 360. Pode observar a presença de um túnel hidrofóbico, com a hélice F" em verde, com os aminoácidos prolina 210, alanina 211, valina 213 e fenilalanina 214 e a alça em vermelho, com os aminoácidos tirosina 457, histidina 458, tirosina 459 e metionina 460.

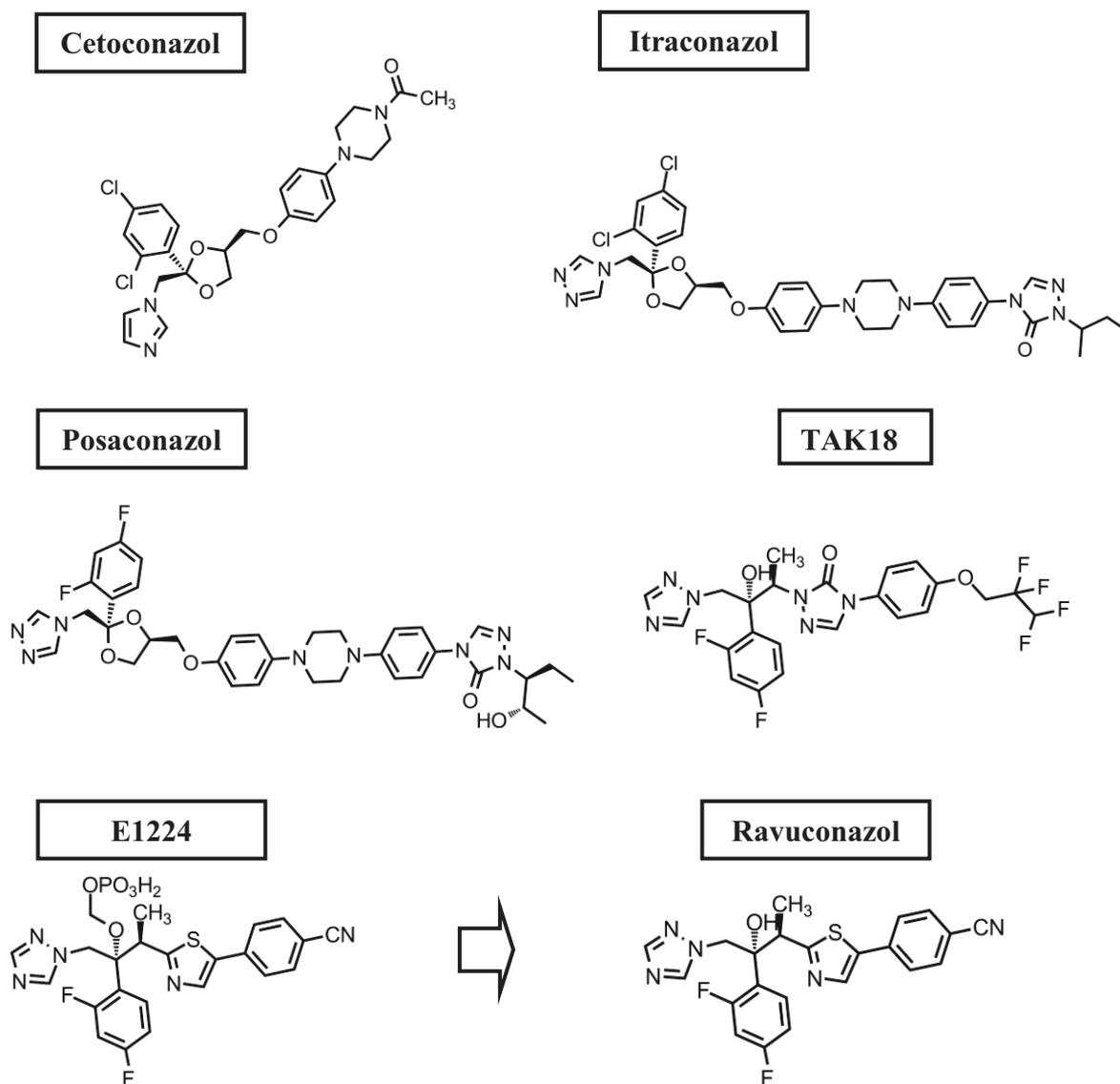


Fonte: Adaptado de CHEN et al, 2010, LEPESHEVA et al, 2010.

Derivados azólicos que agem sobre a enzima lanosterol 14 alfa-desmetilase de fungos demonstraram-se ativos frente à *T. cruzi*. Por exemplo, o posaconazol, um pró-fármaco do ravuconazol (E1224) e TAK mostraram-se ativos contra cepas de *T. cruzi*; cetoconazol e itraconazol (Figura 13) apresentaram grande atividade contra esse parasita *in vitro* (APT, 2003; APT, 1998). Contudo, a dose exigida para eliminar o parasita *in vivo* é maior do que aquela necessária para inibir a síntese de hormônios esteroidais no hospedeiro, podendo causar efeitos hepatotóxicos (AMERY, DE COSTER; CAERS, 1986; MALDONADO, 1993). Entretanto deve-se ressaltar que esses compostos são ativos contra cepas de *T. cruzi* resistentes a

nitrofurano e tem boa atividade na fase crônica da doença de Chagas (URBINA, 2002, URBINA, 2003).

Figura 13. Fármacos que apresentam atividade contra a enzima lanosterol 14 alfa-desmetilase de *T. cruzi* *in vitro*.



Fonte: Adaptado de APT, 2003, SILVA et al, 2012.

Conseqüentemente, essa classe de compostos se mostra promissora para o desenvolvimento de fármacos para o tratamento da fase crônica da doença de Chagas. A fim de atingir esse objetivo pode-se utilizar métodos quimiométricos para extrair o máximo de informação quimicamente relevante a partir de compostos biologicamente ativos.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Utilizar ferramentas computacionais para identificar características físico-químicas e estruturais da classe de inibidores imidazólicos da enzima lanosterol 14 α desmetilase de *T. cruzi*.

2.2 OBJETIVOS ESPECÍFICOS

1. Desenvolver modelos de QSAR baseados em descritores bidimensionais, holograma QSAR e QSAR 3D (CoMFA) robustos e com alto poder preditivo para um conjunto de inibidores imidazólicos da lanosterol 14- α desmetilase (14 α -DM) de *T. cruzi*.
2. Identificar as características físico-químicas e estruturais relevantes para atividade anti-chagásica de derivados de azol.

3. MATERIAIS E MÉTODOS

3.1 CONJUNTO DE DADOS

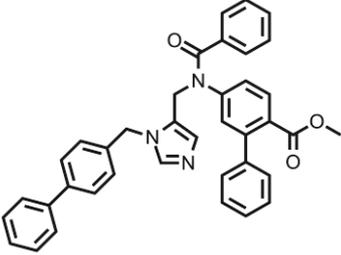
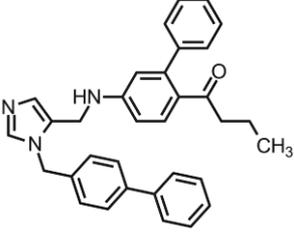
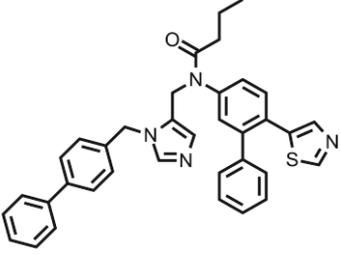
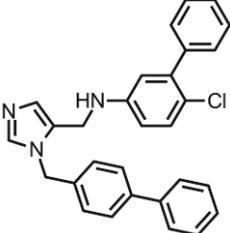
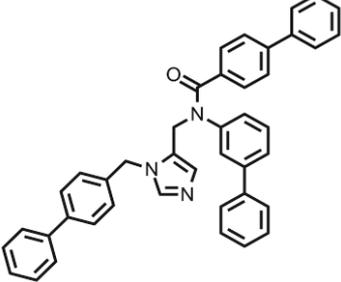
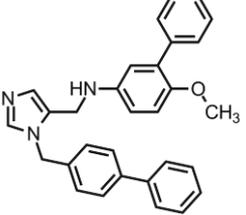
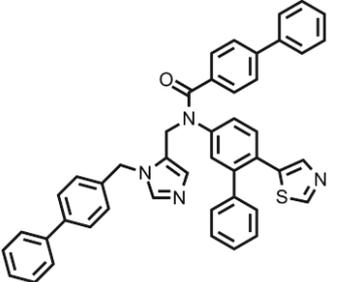
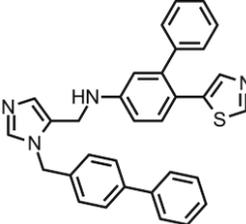
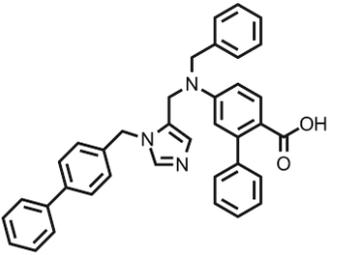
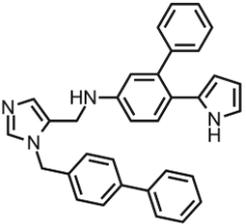
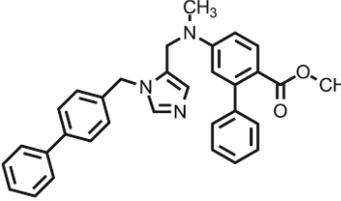
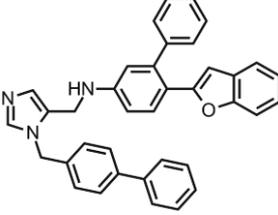
O conjunto de dados é composto por 155 derivados imidazólicos sintetizados e avaliados no laboratório do pelo Dr Michael Gelb (University of Washington, Seattle, Estados Unidos. Parte desses descritores já foram descritos na literatura (SURYADEVARA et al, 2009) e correspondem as moléculas numeradas de 62 a 93 (Tabela 2, com pED₅₀ sem estar em negrito).

Tabela 2. Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
1		7,00	63		8,00
2		6,00	64		7,50
3		6,52	65		7,40

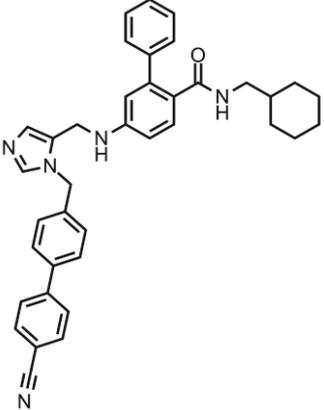
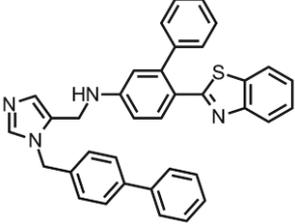
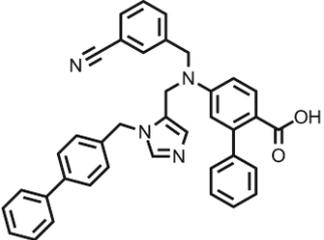
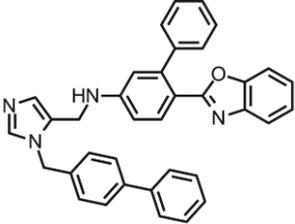
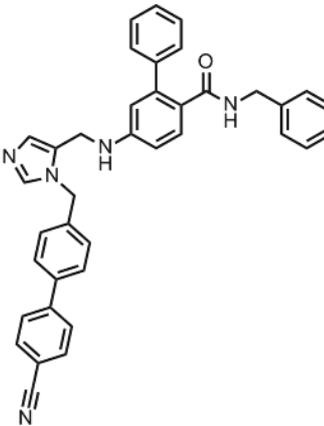
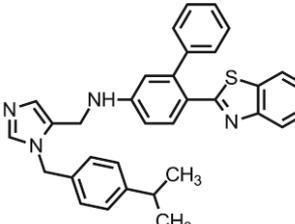
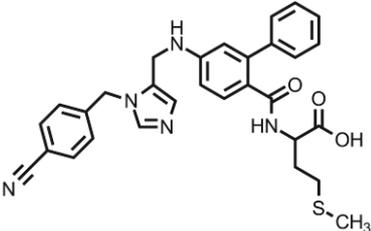
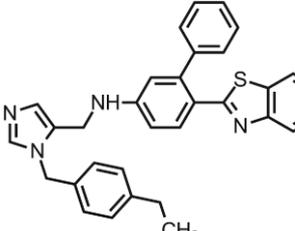
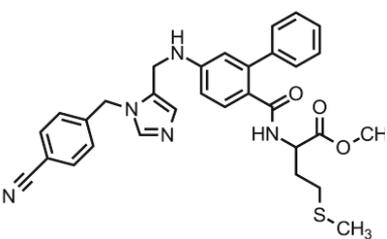
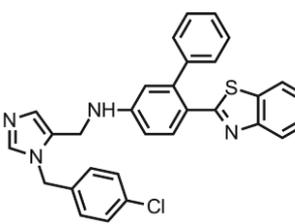
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
4		6,52	66		7,40
5		6,52	67		8,19
6		6,52	68		8,10
7		6,70	69		7,72
8		6,40	70		7,00
9		7,12	71		7,40

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
10		5,52	72		7,82
11		6,70	73		7,00
12		6,22	74		6,15
13		5,52	75		6,82
14		5,70	76		6,60

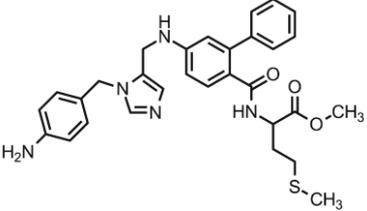
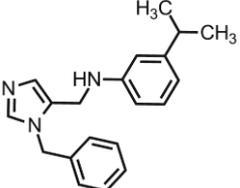
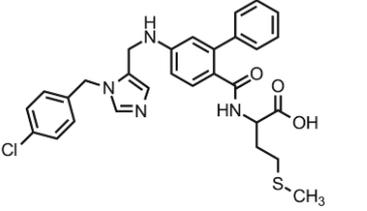
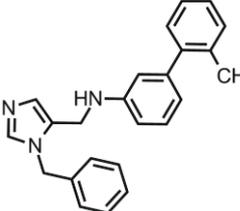
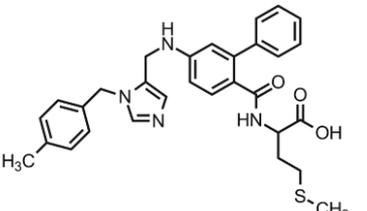
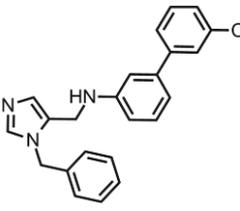
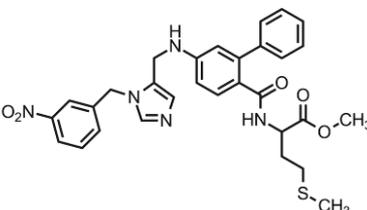
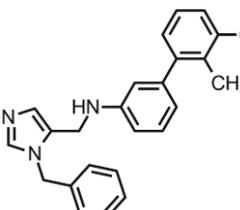
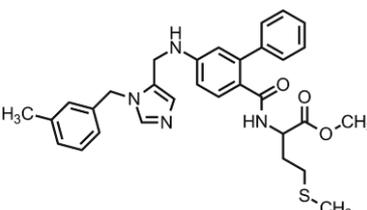
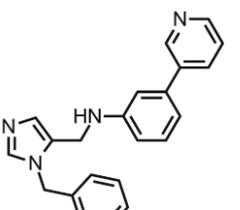
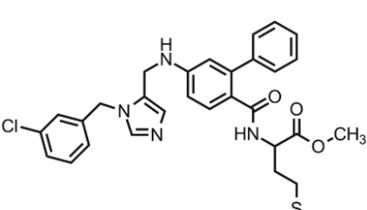
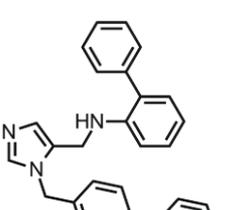
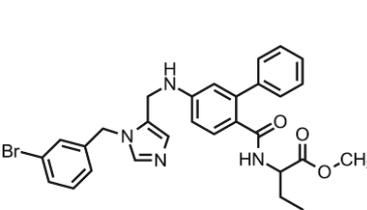
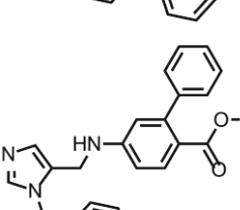
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
15		7,00	77		7,70
16		6,00	78		6,28
17		7,60	79		7,44
18		6,00	80		7,08
19		7,30	81		7,00
20		5,00	82		7,69

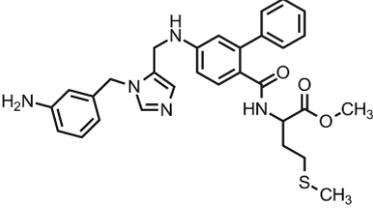
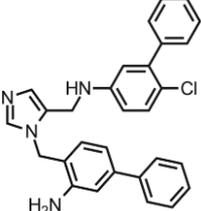
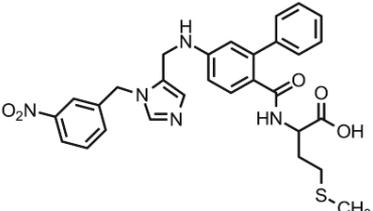
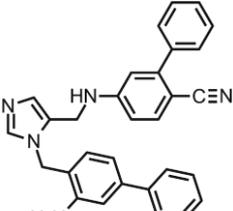
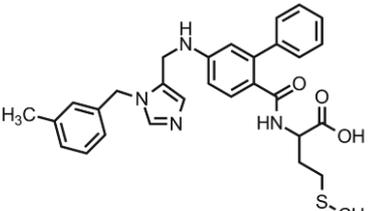
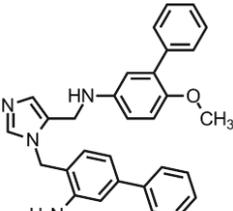
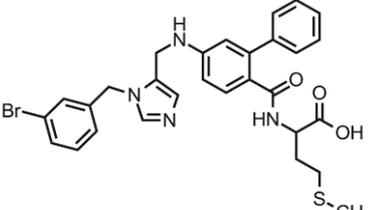
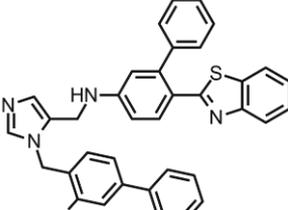
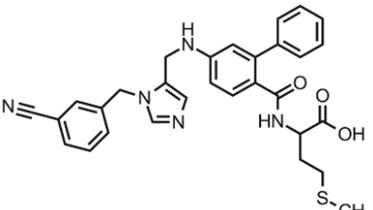
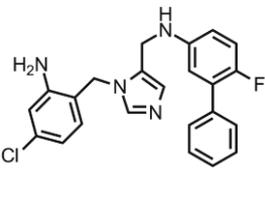
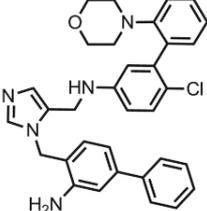
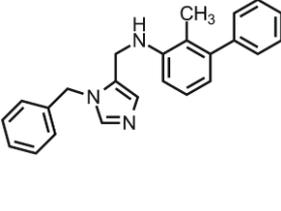
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

Moléculas	pED ₅₀ *	Moléculas	pED ₅₀ *
21 	6,00	83 	6,00
22 	5,00	84 	6,57
23 	5,00	85 	6,26
24 	6,00	86 	6,55
25 	6,00	87 	6,17
26 	6,00	88 	7,10
27 	6,00	89 	7,60

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

Moléculas	pED ₅₀ *	Moléculas	pED ₅₀ *
28 	6,00	90 	8,80
29 	5,00	91 	8,69
30 	5,00	92 	8,84
31 	5,00	93 	9,07
32 	5,00	94 	7,89
33 	7,61	95 	6,63

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
		8,23			7,34
35		7,72	97		7,48
36		6,48	98		7,83
37		8,91	99		7,26
38		8,67	100		8,69
39		6,48	101		7,08
40		7,31	102		6,72

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
41		7,08	103		7,41
42		7,58	104		7,80
43		6,00	105		7,45
44		6,00	106		8,95
45		6,00	107		8,29
46		6,00	108		9,34

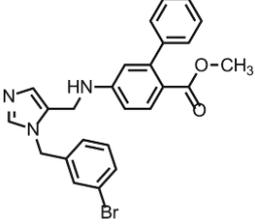
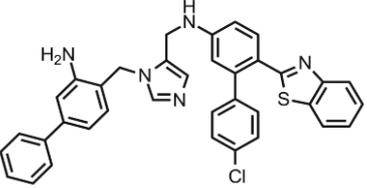
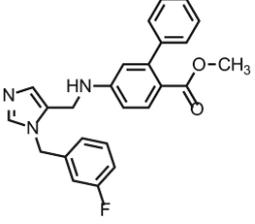
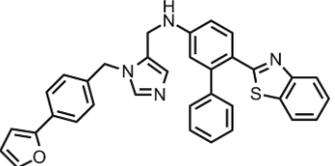
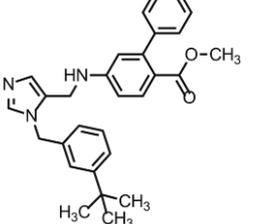
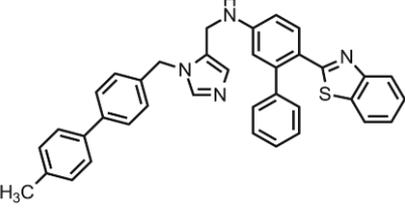
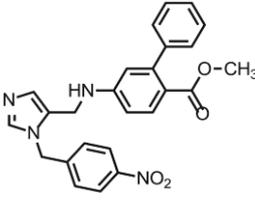
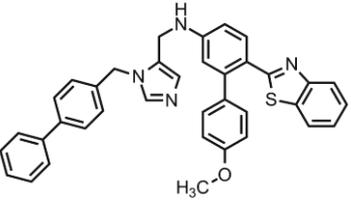
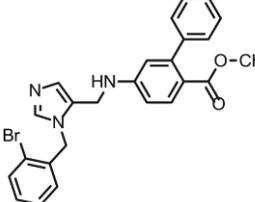
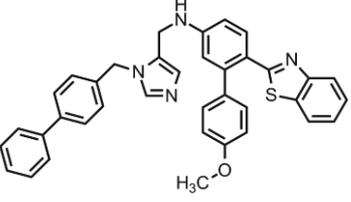
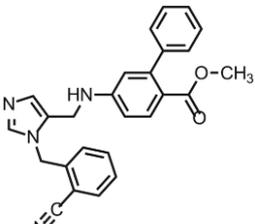
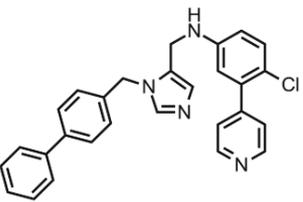
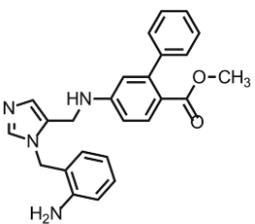
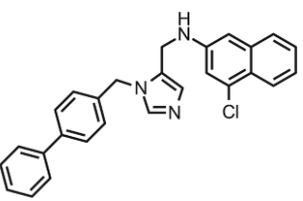
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

Moléculas	pED ₅₀ *	Moléculas	pED ₅₀ *
	7,28		9,35
48	7,34	110	9,45
49	7,82	111	7,84
50	7,10	112	8,04
51	7,64	113	7,15
52	7,52	114	9,10
53	7,70	115	8,23

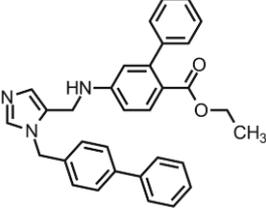
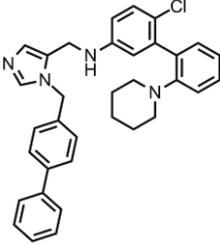
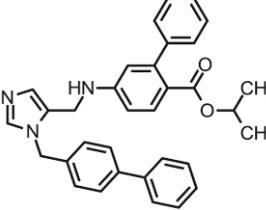
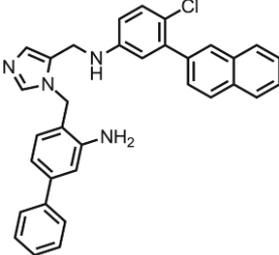
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
54		7,70	116		8,91
55		6,82	117		7,51
56		6,32	118		7,30
57		7,00	119		7,89
58		7,00	120		8,00
59		6,70	121		7,56
60		8,30	122		7,40

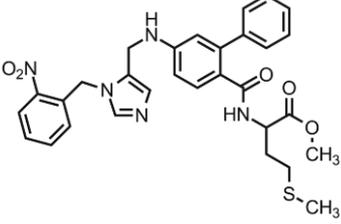
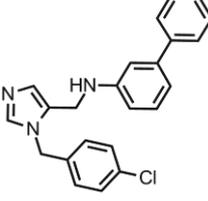
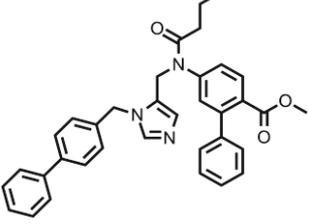
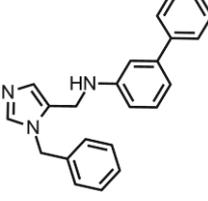
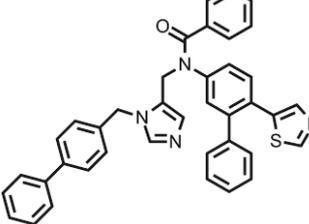
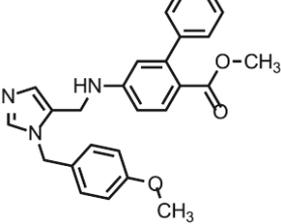
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009).

Tabela 2 (continuação). Moléculas do grupo treinamento

Moléculas	pED ₅₀ *	Moléculas	pED ₅₀ *
61 	8,00	123 	7,68
62 	7,30	124 	8,72

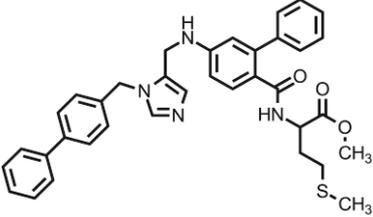
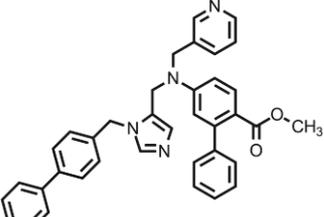
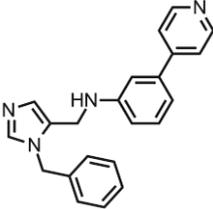
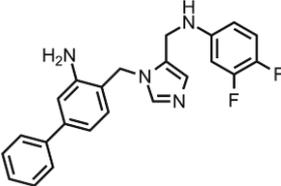
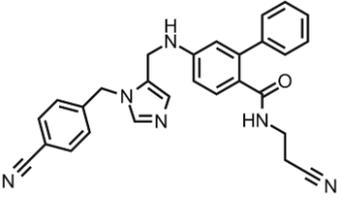
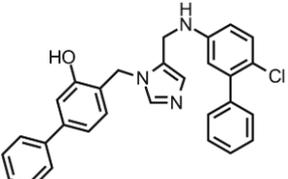
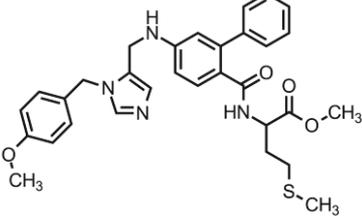
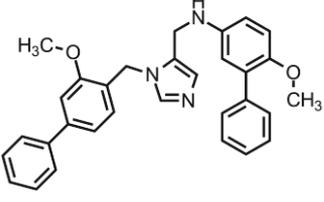
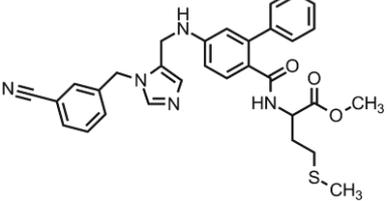
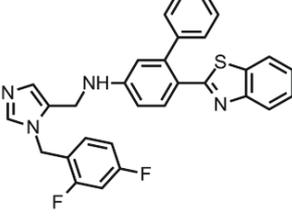
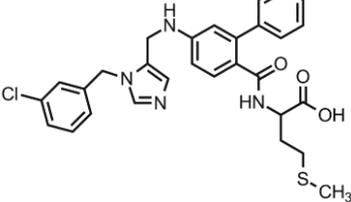
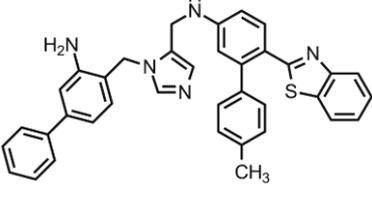
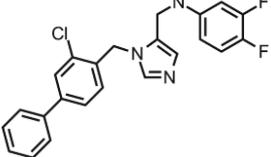
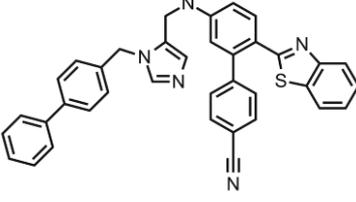
*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 3. Moléculas do grupo teste.

Moléculas	pED ₅₀ *	Moléculas	pED ₅₀ *
1 	6,00	17 	7,10
2 	6,52	18 	6,42
3 	6,12	19 	7,60

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 3 (continuação). Moléculas do grupo teste.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
4		7,00	20		6,70
5		6,29	21		8,49
6		6,00	22		8,12
7		5,80	23		7,90
8		6,00	24		7,68
9		6,12	25		9,03
10		7,42	26		7,47

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Tabela 3 (continuação). Moléculas do grupo teste.

	Moléculas	pED ₅₀ *		Moléculas	pED ₅₀ *
11		7,25	27		6,70
12		7,00	28		8,83
13			29		6,96
14		8,00	30		7,49
15		7,30	31		8,42
16		7,00			

*pED₅₀ em negrito – moléculas não publicadas; pED₅₀ sem negrito – moléculas publicadas (SURYADEVARA et al, 2009)

Os valores de ED₅₀ (concentração necessária para reduzir em 50% a proliferação de *T. cruzi*) que variam de 0,345 a 1000nM foram convertidos em pED₅₀

($-\log ED_{50}$). Posteriormente suas estruturas químicas foram desenhadas em formato 3D e otimizadas através do mecanismo molecular, utilizando o campo de força Tripos, cargas Gasteiger-Huckel, constante dielétrica da água (80) e como parâmetro de convergência um gradiente menor ou igual a 0,001 Kcal/mol. A seguir, empregou-se o método semiempírico PM3 disponível na plataforma SYBYL com as palavras-chave 1SCF XYZ ESP NOINTER SCALE=1.4 NSURF=2 SCINCR=0.4 NOMM para cálculo das cargas atômicas parciais. O conjunto de dados foi dividido em grupo treinamento (124 moléculas, Tabela 2), para a construção do modelo, e teste (31 moléculas, Tabela 3), para fins de validação externa. Os parâmetros utilizados para divisão dos grupos foram diversidade estrutural, como descrito pela Análise de Componentes Principais (PCA) e variação de potência.

3.2 MÉTODOS QUIMIOMÉTRICOS

3.2.1 CÁLCULO E SELEÇÃO DOS DESCRITORES

Os descritores utilizados para a construção dos modelos de QSAR foram calculados com o auxílio do programa DRAGON 5.5. Descritores com valor constante, baixa variância ($< 0,01$) ou com alta correlação entre si ($>0,97$) foram excluídos por não contribuírem com informações relevantes para o desenvolvimento de modelos de QSAR 2D. Na sequência, foram selecionados descritores que pudessem estar correlacionados com a atividade biológica através da geração de modelos preliminares de QSAR obtidos por Regressão Linear Múltipla, utilizando 4 variáveis, como disponível no programa Mobydigs 1.0. Os seguintes parâmetros foram utilizados nessa etapa: 10 populações de 100 indivíduos cada, composto por 4 descritores, evoluindo por 1000 gerações. Cada modelo foi gerado aleatoriamente com taxa de mutação e permutação de 0,5.

A fim de evitar propriedades indesejáveis, tais como alta colinearidade, sobreajuste e presença de variáveis redundantes, a seleção dos indivíduos mais adequados se baseou na função RQK (TODESCHINI, CONSONNI, PAVAN, 2003). Somente variáveis encontradas em modelos de RLM com $q^2 > 0,6$ foram utilizadas nas etapas subsequentes.

3.2.2 ANÁLISE DE COMPONENTES PRINCIPAIS

Os descritores selecionados na etapa anterior foram utilizados no desenvolvimento da Análise de Componentes Principais (PCA, do inglês *Principal Components Analysis*), realizada com o auxílio do programa Pirouette 4.0. O método consiste em condensar a informação original, descrita na forma de matriz, através de outras duas outras matrizes mais simples; uma matriz de pontuação que representa as amostras num novo eixo de coordenadas, denominadas PCs, e uma matriz de peso, que descreve a contribuição de cada variável original para as PCs, que são ortogonais entre si. Primeiramente, os descritores foram autoescalados para eliminar problemas de unidade e variância. A seguir critérios de validação-cruzada e variância explicada foram empregados para determinar o número ótimo de Componentes Principais (PCs). Os resultados obtidos foram analisados através do gráfico de pontuação (*score plot*).

3.2.3 QSAR BASEADO EM DESCRITORES BIDIMENSIONAIS

Os descritores físico-químicos, selecionados com o auxílio do programa Mobydigs, foram autoescalados e usados como variáveis independentes, enquanto a atividade biológica (pED₅₀) foi utilizada como variável dependente nas análises de PLS realizadas no programa Pirouette 4.0. O número ótimos de PCs foi determinado por validação cruzada deixo um para fora (*LOO* do inglês *leave-one-out*) (q^2). Durante a otimização dos modelos, descritores com contribuições discretas para o vetor de regressão foram paulatinamente eliminados desde que isso não acarretasse em redução dos valores de q^2 . A consistência interna dos modelos foi avaliada através dos valores de r^2 e q^2 do grupo treinamento (Equação 1 e 2 respectivamente) (CONSONNI, BALLABIO, TODESCHINI, 2009).

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y}_i)^2}{\sum (Y_i - \bar{Y}_i)^2} \quad \text{Equação 1}$$

$$q^2 = 1 - \frac{\sum (\hat{Y}_i - \bar{Y}_i)^2}{\sum (Y_i - \bar{Y}_i)^2} \quad \text{Equação 2}$$

onde Y_i é o valor da atividade experimental; \hat{Y}_i é o valor da atividade predita; \bar{Y}_i a média dos valores experimentais (LEACH, GILLET, 2007).

3.2.4 HOLOGRAMA QSAR (HQSAR)

A construção e análise dos modelos de HQSAR foram realizados no programa Sybyl-X1.1, executado no sistema operacional Windows. Durante o processo de construção modelo, foram selecionadas 12 séries de comprimento do holograma (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, e 401). Várias combinações de distinção de fragmentos (Átomo (A), Ligação (B), Conectividade (C), Hidrogênio (H), Quiralidade (Ch) e Doador e Aceptor de Hidrogênio (DA)) foram selecionados aleatoriamente durante o desenvolvimento do modelo de HQSAR, considerando inicialmente o tamanho padrão de fragmento (4-7 átomos). Posteriormente, foram gerados modelos utilizando tamanhos de fragmentos distintos (3-5; 5-8; 6-9) para o melhor modelo obtido até esse momento. Os hologramas foram usados como variáveis independentes durante a análise de PLS, na qual o número ótimo de PCs foi determinado por validação cruzada do tipo LOO. A consistência interna dos modelos foi avaliada através dos valores de q^2 enquanto o ajuste do modelo foi medido através do valor de r^2 .

3.2.5 ANÁLISE COMPARATIVA DOS CAMPOS MOLECULARES (COMFA)

Estudos de QSAR 3D que utilizam descritores de campo são profundamente influenciados pelo alinhamento das moléculas no espaço, por essa razão 3 estratégias de alinhamento (acoplamento molecular, similaridade morfológica e máxima estrutura em comum) foram empregados nos estudos de CoMFA. Maiores detalhes sobre cada uma dessas estratégias segue abaixo

3.2.5.1 ALINHAMENTO POR ACOPLAMENTO MOLECULAR

A estrutura cristalográfica da lanosterol 14 alfa-desmetilase empregada para realizar o acoplamento molecular (código PDB 3K1O, resolução de 2,89 Å) passou pelas seguintes etapas de preparação: remoção do ligante presente no sítio ativo enzima (posaconazol); eliminação das moléculas de água; introdução dos átomos de hidrogênio e análise individual do estado de protonação e tautômeros dos resíduos

de histidina e glutamina encontrados no sítio ativo da enzima. Todas essas etapas foram realizadas no módulo Biopolymer no programa SYBYL-X1.1. A seguir, o espaço de busca para as simulações foi definido como uma esfera com 12 Å de raio a partir do carbono central do posaconazol encontrado na estrutura cristalográfica ($x = -43,271$; $y = 9,025$; $z = -17,282$), disponível no programa GOLD 4.1.2. Para a geração das conformações, foi considerado a flexibilidade dos ligantes, como o nitrogênio piramidal, amidas e conformações de moléculas cíclicas. Também foram utilizados parâmetros de algoritmo genético para gerar conformações mais fidedignas. Durante a evolução das populações, de tamanho 100, foi utilizado 4 números de ilhas para que diversas subpopulações de cromossomos fossem criadas. Os operadores genéticos incluem 10 migrações de cromossomos individuais, 95 permutações e 95 mutações. Empregou-se a técnica de nicho para manter a diversidade conformacional dentro da análise. Durante a execução dessa técnica, adiciona-se um novo indivíduo na população, sendo que, se houver mais que 2 indivíduos dentro do nicho, o indivíduo novo substitui o pior membro (CHEN et al, 2006). Somente as poses de maior pontuação, segundo a função de pontuação Goldscore P450 (KIRTON et al, 2005) foram consideradas para a construção dos modelos CoMFA.

3.2.5.2 ALINHAMENTO POR SIMILARIDADE MOLECULAR

O alinhamento realizado com o auxílio do programa Surfex-Sim, disponível na plataforma Sybyl-X1.1, emprega uma função de similaridade morfológica para alinhar um conjunto de moléculas. Nessa abordagem, foram realizados dois alinhamentos: sobre um ligante conhecido (no caso o posaconazol) e sobre os 5 ligantes mais potentes do conjunto treinamento (vide Tabela 2), considerando a flexibilidade rotacional. Durante essa etapa, cada molécula foi dividida em até 100 fragmentos para os quais foram gerados até 20 conformações diferentes. Os fragmentos, após passarem por um processo de minimização de energia, foram utilizados para criar, de forma incremental, o alinhamento espacial das moléculas como base nos seus grupos hidrofóbicos e doadores/aceptores de ligação de hidrogênio, além do seu volume e forma. Após cada etapa de construção incremental das moléculas, nova etapa de minimização de energia foi realizada e conformações com o desvio médio da raiz quadrada (RMSD, do inglês *root mean*

square deviation) inferior a 0,05 foram agrupadas. A solução de maior pontuação foi utilizada como molde (rígido) para o alinhamento das demais moléculas do grupo treinamento e teste, usando os mesmos parâmetros descritos anteriormente.

3.2.5.3 ALINHAMENTO PELA MÁXIMA SUBESTRUTURA COMUM

O alinhamento por máxima estrutura comum é um método que se baseia na premissa de que o esqueleto comum de um conjunto de dados contribui igualmente para a atividade biológica e pode ser considerada como uma constante (YUAN et al, 1990, CHAUDHAERY, ROY, SAXENA, 2009). Nesse sentido, o esqueleto principal comum a todas as estruturas é utilizado como base para sobreposição, sendo que o melhor alinhamento é obtido através da otimização da raiz quadrada média (RMS) do átomo/centro dos ligantes (CHAUDHAERY, ROY, SAXENA, 2009, ZHUO et al, 2008). A energia conformacional das moléculas do conjunto foram obtidas em um mínimo energético, pois acredita-se que essa conformação se aproxima da conformação bioativa (KUBINYI, 1998).

Após o alinhamento, as propriedades estéricas e eletrostáticas foram calculadas de acordo com os potenciais de Lennard-Jones e de Coulomb (Equação 3 e 4 respectivamente)

$$E_{vdw} = \sum_{i=1}^n \left(\frac{1}{A_{ij} r_{ij}^{12}} - \frac{1}{C_{ij} r_{ij}^6} \right) \quad \text{Equação 3}$$

$$E_C = \sum_{i=1}^n \frac{q_i q_j}{D r_{ij}} \quad \text{Equação 4}$$

As moléculas do conjunto treinamento, alinhadas espacialmente conforme descrito anteriormente, foram circundadas por uma grade tridimensional que se estende pelo menos 4Å além do último átomo de cada molécula e com espaçamento de 2,0 Å entre seus pontos de intersecção. A seguir, campos estereoquímicos e eletrostáticos de interação entre as moléculas e uma sonda (carbono sp³ com carga +1) foram calculados em cada ponto da grade, utilizando parâmetros padrão do campo de força Tripos e um valor de ponto de corte de 30kcal/mol. Os valores obtidos foram então correlacionados com a atividade biológica das moléculas pelo método dos Mínimos Quadrados Parciais (PLS) (ZHENG et al, 2011, HONÓRIO et

al, 2007). Num segundo momento, a influência do espaçamento da grade (0,5, 1,0, e 1,5) e do ajuste dos campos pelo coeficiente desvio padrão (*SDC*, do inglês *standard deviation coefficient*) sobre os parâmetros estatísticos do modelo também foram avaliados.

3.2.6 VALIDAÇÃO DOS MODELOS

O ajuste dos modelos de QSAR foi avaliado através do quadrado do coeficiente de correlação (r^2), enquanto o valor de q^2 (obtido por validação cruzada) (LEACH, GILLET, 2007) foi usado para estimar a consistência interna dos modelos e determinar o número ótimo de componentes principais que deveria ser utilizado nas regressões por PLS.

A capacidade preditiva dos modelos de QSAR (r^2_{pred}) foi investigada através de métodos de validação externa utilizando um conjunto de 31 moléculas não utilizadas na calibração dos modelos de QSAR. Existem três maneiras de se calcular o r^2_{pred} (CHIRICO, GRAMATICA, 2011). Nesse trabalho esse valor foi calculado como mostrado nas três equações abaixo:

$$r^2_{predF1} = 1 - \frac{\sum_{i=1}^{n_{TESTE}} (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{n_{TESTE}} (Y_i - \bar{Y}_{TR})^2} \quad \text{Equação 5}$$

onde o somatório refere-se aos dados do grupo teste, com \hat{Y}_i representa a atividade biológica predita do conjunto teste; Y_i a atividade biológica experimental do grupo teste \bar{Y}_{TR} a média da atividade biológica do grupo treinamento. O uso da média dos valores do conjunto treinamento está relacionado ao controle da distância entre os dois conjuntos (treinamento e teste).

Uma proposta alternativa foi por Schuurmann e colaboradores (2008) para o cálculo de r^2_{pred} (equação 6). Essa alternativa não considera informações sobre o conjunto treinamento.

$$r^2_{predF2} = 1 - \frac{\sum_{i=1}^{n_{TESTE}} (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{n_{TESTE}} (Y_i - \bar{Y}_{TESTE})^2} \quad \text{Equação 6}$$

onde se tem a somatória dos elementos do grupo teste, onde \hat{Y}_i representa a atividade biológica predita do conjunto de dados teste; Y_i a atividade biológica experimental do grupo teste e \bar{Y}_{TESTE} a média da atividade biológica do grupo teste.

Após a formulação de conceitos da capacidade preditiva do conjunto teste com os valores de r^2_{predF1} e r^2_{predF2} , a utilização do grupo treinamento para prever as moléculas do grupo externo foi bastante discutida na comunidade científica. Segundo alguns autores, a informação a cerca do grupo treinamento é importantíssimo para a predição enquanto que outros autores consideram uma desvantagem. No intuito de superar essas discussões, outra maneira de cálculo da capacidade preditiva dos modelos foi criada. (Equação 7);

$$r^2_{predF3} = 1 - \frac{\left[\sum_{i=1}^{n_{TESTE}} (\hat{Y}_i - Y_i)^2 \right] / n_{TESTE}}{\left[\sum_{i=1}^{n_{TR}} (Y_i - \bar{Y}_{TR})^2 \right] / n_{TR}} \quad \text{Equação 7}$$

onde se tem a somatória dos elementos do grupo teste no numerador e a somatória do grupo treino no denominador, onde \hat{Y}_i representa a atividade biológica predita do conjunto de dados teste; Y_i a atividade biológica experimental do grupo teste (numerador) e treinamento (denominador) e \bar{Y}_{TR} a média da atividade biológica do grupo treinamento.

A fim de avaliar a precisão do modelo a modificações no conjunto treinamento, foi realizado o procedimento de reamostragem (do inglês *bootstrapping*) e o método de retirada (LMO, do inglês *leave-many-out*). No procedimento de reamostragem, as moléculas são divididas aleatoriamente em grupos, que podem term compostos repetidos, a partir dos quais novos modelos de QSAR foram gerados. Os valores de q^2 e r^2 obtidos nesse processo foram então comparados com aqueles do modelo original. Devido a natureza estocástica desse procedimento ele foi executado 5, 10, 50 e 100 vezes. O LMO, retirou um conjunto de 5, 10, 15 e 20 compostos aleatoriamente e o modelo é refeito (OTTO, 2007, CONSONNI, BALLABIO, TODESCHINI, 2009; ANDRADE et al, 2008). A robustez dos modelos de CoMFA foi avaliado pelo cálculo de q^2 e teste T.

4. RESULTADOS E DISCUSSÃO

Posaconazol, E1224 (profármaco que se converte em ravuconazol, desenvolvido por Eisai, Japão) e TAK187 (desenvolvido por Takeda, Japão) (Figura 13) são derivados triazólicos que têm se mostrado promissores como antichagásicos e estão em estudo de fase clínica II e I respectivamente (CLAYTON, 2010, SILVA et al, 2012). Entretanto, derivados triazólicos apresentam algumas desvantagens (ex. problemas gastrointestinais graves e alto custo financeiro) que podem limitar sua utilização no tratamento de longo prazo da fase crônica da doença. O cetoconazol e itraconazol (Figura 13) também demonstraram atividade contra cepas de *T. cruzi* e são mais baratos que os fármacos anteriores, no entanto apresentam sérios efeitos colaterais. Na tentativa de contornar parte das limitações encontradas pode-se recorrer a síntese de compostos com menor complexidade e portanto, mais baratos (SURYADEVARA et al, 2009, LEPESHEVA, 2010).

Seguindo esse raciocínio, o grupo liderado pelo Dr. Michael Gelb (University of Washington, Seattle, Estados Unidos) sintetizou um conjunto de 75 derivados dialquilimidazóis que foram avaliados quanto a sua capacidade de inibir a proliferação de *T. cruzi* (SURYADEVARA et al, 2009). Visando identificar o alvo macromolecular dessas moléculas e considerando a semelhança estrutural delas com inibidores de lanosterol 14 alfa-desmetilase de fungos, foram realizados teste de inibição cinética da enzima lanosterol 14 alfa-desmetilase de *T. cruzi*. O deslocamento do pico de absorção do ferro do grupo heme para 432-410 observado nesses ensaios é característico de moléculas que interagem diretamente com o átomo de ferro, provando assim que essa classe de moléculas inibe a via de síntese de ergosterol no parasita (SURYADEVARA et al, 2009, BUCKNER, 2003). Em paralelo, visando demonstrar que as propriedades farmacocinéticas dos compostos mais potentes são favoráveis a atividade farmacológica desejada, decidiu-se continuar empregando modelos animais. Os esforços sintéticos desenvolvidos no laboratório do Dr. Gelb levaram a síntese de 155 moléculas cuja dose efetiva para eliminar os parasitas em 50% das células tratadas (ED_{50}) varia de 0,345 nM a 1000 nM (Tabela 2). Embora parte do planejamento dessas moléculas tenha sido guiado pelos resultados biológicos empregados e por dados estruturais, até o presente momento nenhuma abordagem quantitativa foi empregada para investigar as exigências estereoquímicas e eletrônicas que determinam sua atividade biológica

frente ao *T. cruzi*. Na tentativa de reduzir essa lacuna no conhecimento acerca das relações entre a estrutura química e a atividade biológica dessa classe de moléculas, modelos de QSAR 2D e 3D foram desenvolvidos.

A atividade biológica das moléculas em estudo foi obtida em condições experimentais padronizadas, minimizando assim erros pré-analíticos de medida da atividade e variações interlaboratoriais, o que torna o conjunto de dados adequado para construção de modelos confiáveis de QSAR (SCIOR et al, 2009). O conjunto de compostos apresenta grande diversidade química (Tabela 2 e 3), embora tenha uma estrutura química em comum (ver Tabela 2). O anel imidazol, presente em todas as moléculas, assume papel fundamental no processo de reconhecimento molecular frente a enzima lanosterol 14 alfa-desmetilase por interagir com o ferro do grupo heme (CHE, 2009, PODUST, POULOS, WATERMAN, 2001).

Antes da geração dos modelos, o conjunto de dados foi dividido em grupo treinamento e grupo teste. O grupo treinamento é aquele destinado a construção e calibração dos modelos de QSAR enquanto que o grupo teste é utilizado na avaliação externa do modelo. A escolha adequada do conjunto treinamento e teste é fundamental para o desenvolvimento de modelos de QSAR robustos e preditivos (LEONARD, ROY, 2006). Portanto, foram adotadas estratégias para assegurar que a diversidade química total e a variação da atividade biológica estivessem representadas nas moléculas do conjunto teste. A fim de contemplar a primeira questão, a diversidade química foi representada através de descritores topológicos e físicoquímicos bidimensionais disponíveis no programa DRAGON 5.5. Entretanto, devido ao grande número de descritores calculados (307), decidiu-se implementar uma etapa de pré-seleção, através da construção de modelos de regressão linear múltipla com até quatro variáveis, como disponível no programa MobyDigs 1.0. Maiores detalhes sobre regressão linear múltipla serão abordados adiante. Por hora, basta mencionar que nessa etapa todas as moléculas do conjunto de dados foram utilizadas e que essa metodologia permitiu a seleção de um subconjunto de 53 descritores que tiveram sua informação condensada através de PCA. Isso permite uma análise visual da diversidade química das moléculas num espaço de poucas dimensões (GIACOMINO, 2011).

A PCA é classificada como um método não supervisionado, visto que a atividade biológica não é levada em consideração durante a análise. A idéia chave da PCA é condensar a informação original, representando as amostras em um novo

eixo de coordenadas, denominadas PCs (OTTO, 2007). A grande vantagem dessa metodologia consiste em reduzir a dimensionalidade de um conjunto de dados, mantendo o máximo possível das informações presentes nos dados originais nas (PCs) (GIACOMINO, 2011) e eliminar problemas de colinearidade entre os descritores, uma vez que as PCs são ortogonais entre si (WOLD, ESBENSEN, GELADI, 1987).

Como as PCs são determinadas com base do critério de variância máxima, a primeira PC contém a maior variância dos dados, seguida pela segunda e assim sucessivamente (OTTO, 2007). Portanto, um pequeno número de PCs consegue descrever grande parte da informação original. Neste trabalho, utilizamos duas estratégias para determinar o número ótimo de PCs: 1) validação cruzada, onde os descritores de um composto são removidos do grupo de dados e se calcula o número de PCs necessárias para descrever a informação contida nos descritores que foram excluídos do modelo. Esse cálculo é repetido até que todos os conjuntos de descritores tenham sido excluídos pelo menos uma vez (LOO); 2) variância explicada, segundo o qual componentes que descrevem $\leq 5\%$ da variância total não devem ter grande influência na descrição das moléculas e portanto podem ser desconsideradas. A partir desses critérios o número ótimo de PCs seria cinco (Tabela 4). Entretanto, apenas as duas primeiras PCs, que descrevem 59,03% da informação original, foram consideradas para análise gráfica dos resultados.

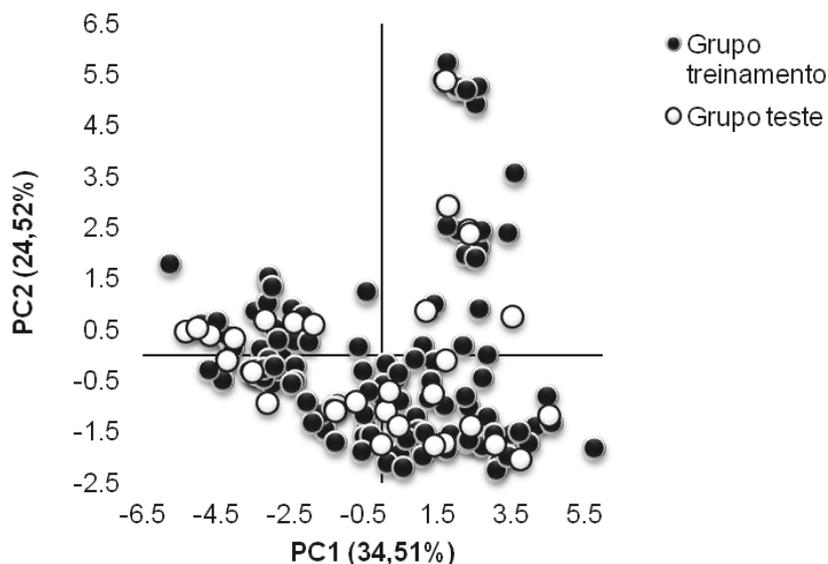
Tabela 4. Valores da variância explicada e acumulada nas PCs do modelo de PCA.

Número de PCs	Variância explicada	Variância acumulada
PC1	34,51	34,51
PC2	24,52	59,03
PC3	8,48	67,08
PC4	5,70	72,78
PC5	5,37	78,14
PC6	3,31	81,45

A visualização das moléculas no espaço químico definido por PC1 e PC2 (Figura 14) revela que estão distribuídas ao longo dos 4 quadrantes. Para representar a diversidade química estrutural do conjunto, seria necessário selecionar de forma aleatória representantes de todos os quadrantes. Dessa forma, 20% das

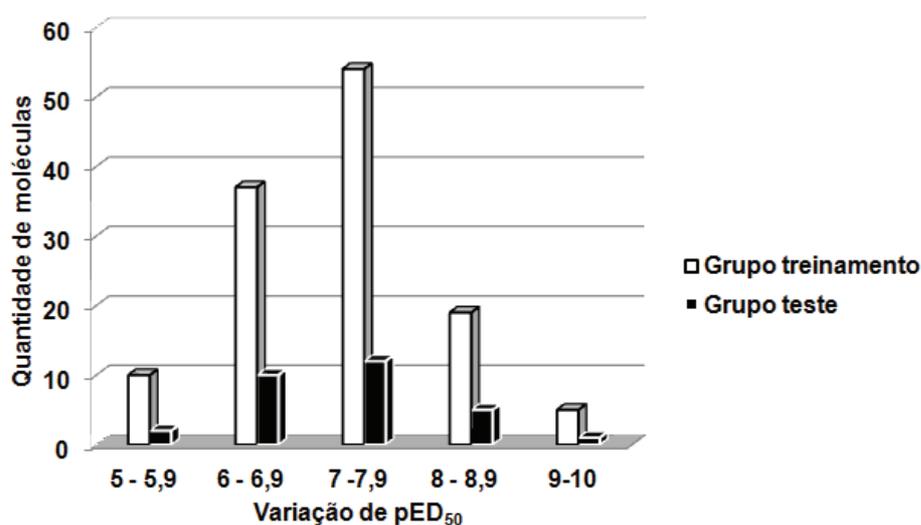
moléculas do conjunto total foram selecionadas para compor o conjunto teste. Outra informação obtida a partir desse gráfico é que nenhuma das amostras apresenta valores de PC1 e PC2 muito diferentes das demais moléculas, sugerindo que não existem outliers no conjunto de dados.

Figura 14. Distribuição dos derivados azólicos no espaço químico definido por PC1 e PC2. Para a divisão do grupo teste optou-se por escolher moléculas que fazem parte de todos os quadrantes.



Atenção especial foi exercida para garantir que esses compostos também apresentassem um perfil de variação de atividade biológica similar aquele encontrado no conjunto treinamento (Figura 15).

Figura 15. Distribuição dos derivados azólicos no grupo treino (em preto) e teste (em vermelho), segundo sua atividade biológica.



Vários métodos matemáticos têm sido usados para desenvolver modelos de QSAR. O método de regressão linear múltipla é um exemplo desses métodos no qual tem como objetivo encontrar valores para os coeficientes de ajuste m e para a constante c que minimizam a soma das diferenças entre os valores de atividade biológica preditos pela equação e aqueles observados experimentalmente (LEACH, GILLET, 2007). Portanto, o modelo de regressão foi desenvolvido com o auxílio do programa MOBIDIGS. Os valores estatísticos derivados desse modelo foram satisfatórios ($q^2=0,70$; $r^2=0,72$), considerando que valores de q^2 e r^2 devem ultrapassar 0,5 e a diferença entre eles não deve ser maior que 0,3 para evitar sobreajuste (OECD, 2007).

Embora a validação interna exclua uma molécula do conjunto para validação do modelo, durante o procedimento todas as moléculas são incluídas no processo global, ou seja, todas as informações são usadas (CHIRICO, GRAMATICA, 2011). Para a avaliação da capacidade preditiva de moléculas congênicas não utilizadas na construção do modelo, torna-se necessário realizar procedimentos de validação externa (CHIRICO, GRAMATICA, 2011, SCHUURMANN et al, 2008, CONSONNI, BALLABIO, TODESCHINI, 2009). A fim de avaliar o poder preditivo do modelo, três estratégias de aferição de r^2_{pred} (Equação 5, 6 e 7) foram utilizadas. Na primeira estratégia (Equação 5) a capacidade preditiva foi calculada de forma parecida com q^2_{LOO} , com a diferença que os valores dos dados do grupo teste e a média do grupo treinamento são usados nos cálculos. A escolha de usar a média do grupo treinamento parte da premissa da obtenção de um valor de referência para a média independente da composição do grupo teste. Essa fórmula permite manter o controle da distância entre o grupo treinamento e o grupo teste, sendo que um bom modelo é aquele no qual o grupo teste alcança o domínio de resposta biológica do grupo treinamento (CHIRICO, GRAMATICA, 2011, CONSONNI, BALLABIO, TODESCHINI, 2009). A avaliação do poder preditivo externo da Equação 5, considerando a média do grupo treinamento proporcionou valores de 0,58 (Figura 17)

Entretanto, a Equação 5 pode superestimar ou subestimar a capacidade de predição em determinadas condições, sendo esse comportamento inadequado para a análise. Superestima quando os valores de atividade biológica dos elementos do grupo teste estão no limite comparados ao conjunto treinamento e subestima quando os valores estão centrais. Em outras palavras, essa desvantagem acontece

quando os valores de atividade biológica dos constituintes do grupo teste não representam a distribuição do grupo treinamento (CHIRICO, GRAMATICA, 2011).

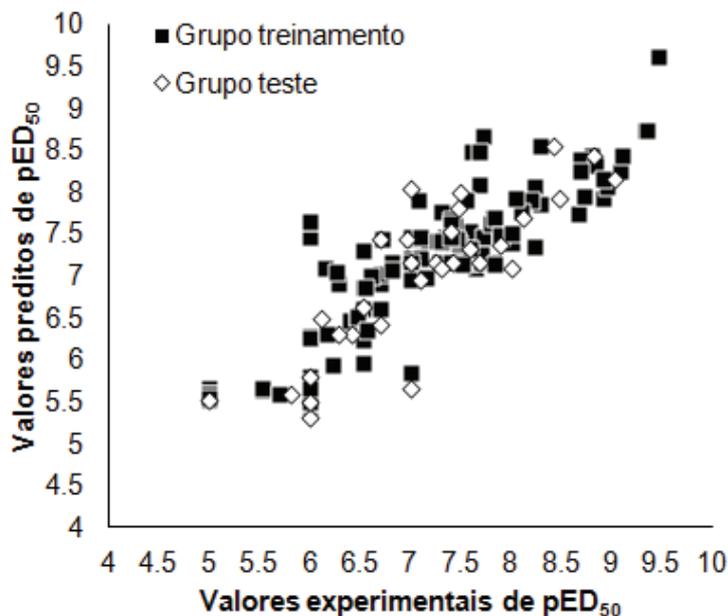
Uma outra maneira de se calcular o poder preditivo externo foi elaborado por Schuurman (2008) (Equação 6), que formulou a hipótese de que o uso da média do grupo teste ao invés do grupo treinamento fornecem estimativas mais confiáveis da capacidade preditiva externa. (CONSONNI, BALLABIO, TODESCHINI, 2009, SCHUURMANN et al, 2008). Esse cálculo é uma forma parecida q^2 LOO, mas para o grupo teste, ou seja, ele avalia o poder preditivo desse grupo. No entanto, esse cálculo não apresenta informações sobre o modelo de validação cruzada no qual está sendo analisado o grupo teste, o que não é uma idéia muito viável. Isso porque essa informação pode mudar constantemente a depender da composição deste grupo (CONSONNI, BALLABIO, TODESCHINI, 2009, CHIRICO, GRAMATICA, 2011). Por isso, é necessário um grupo teste com diversidade química e de atividade biológica. A análise do poder preditivo do grupo teste realizada através da Equação 6 possibilitou o valor de $r^2_{pred} = 0,58$. Como pode ser observado, os valores do poder preditivo do grupo teste da Equação 5 e 6 são iguais, ou seja, o grupo teste está com poder preditivo considerável e não está sendo superestimado ou subestimado quando considera-se o grupo treinamento. Isso se deve a um grupo teste com boa variabilidade química e biológica, uniformemente distribuída no grupo treinamento.

Uma outra alternativa para o cálculo do poder preditivo externo (Equação 7) pode ser considerada, onde informações do grupo treinamento é avaliada no denominador e do grupo teste no numerador e ambos numerador e denominador são divididos pelo número dos elementos correspondentes (CHIRICO, GRAMATICA, 2011). Entretanto, essa equação é independente da distribuição do grupo teste e por isso não será abordada no estudo.

Portanto, os resultados dos modelos foram avaliados levando em consideração as duas maneiras de calcular a predição externa para validação dos modelos. O modelo desenvolvido pela metodologia de descritores bidimensionais apresentou poder preditivo insatisfatório, já que o valor do r^2_{pred} se distancia do q^2 tanto na Equação 5 ($r^2_{pred} = 0,58$) quanto na Equação 6 ($r^2_{pred} = 0,58$) (Figura 16). Esse resultado permite concluir que as moléculas de fora não possuem poder preditivo similar as moléculas do grupo treinamento. Logo, o problema pode ser decorrente do método de RLM, que se mostra sensível a presença de colinearidade entre as variáveis, proporcionando um sobreajuste no modelo, reduzindo sua

estabilidade e dificultando a análise da relação da atividade molecular de acordo com o espaço químico amostral do conjunto (SLINKER, GLANTZ, 2008). Portanto, a geração de modelos por RLM foi usada apenas como técnica para seleção de variáveis.

Figura 16. Valores reais versus valores preditos de pED_{50} de derivados imidazólicos inibidores da enzima lanosterol 14 alfa-desmetilase de *T. cruzi* no modelo de RLM.

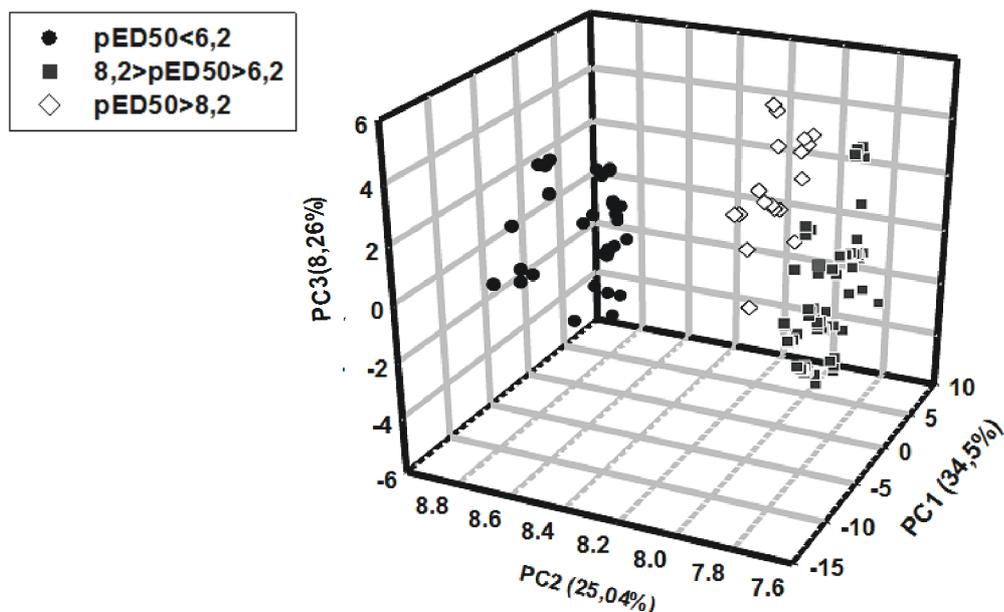


Para contornar tal problema, foi utilizado a técnica de PCA (página 64) como análise exploratória dos dados, permitindo identificar agrupamentos naturais entre as moléculas. A Análise de Componentes Principais se baseia na projeção das variáveis em uma nova perspectiva onde as informações relevantes se sobressaem, ou seja, as variáveis correlacionadas são agrupadas em uma nova variável (PCs), sendo mutuamente ortogonais (FERREIRA, MONTANARI, GAUDIO, 2002). Assim, a visualização das propriedades moleculares está por conta dos descritores considerados no modelo. A fim de visualizar se os descritores oriundos do modelo de RLM são capazes de descrever as propriedades moleculares do conjunto de dados, o método de PCA foi realizado. Os descritores foram autoescalados para essa etapa.

Ao se analisar o gráfico de pontuação, verifica-se que PC1 descreve 34,36% da informação original, PC2 15,97%, PC3 10,28%, totalizando 60,61% da variância total. Entretanto, a PC1 não é capaz de separar as moléculas de acordo com sua atividade molecular. De fato, a subdivisão das moléculas com alta, média e baixa

atividade biológica só se torna clara quando as 3 primeiras PCS são consideradas (Figura 17).

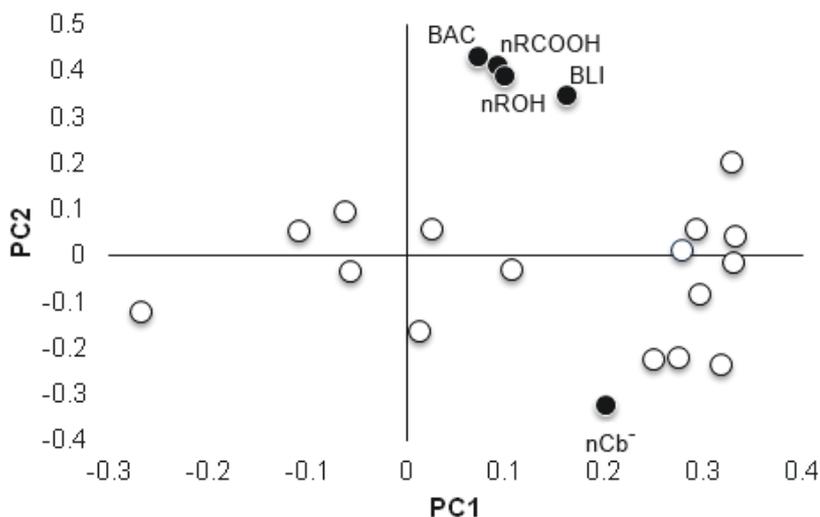
Figura 17. Gráfico de pontuação do modelo de PCA para os derivados azólicos ativos contra a enzima lanosterol 14 alfa-desmetilase de *T. cruzi*.



Esses dados mostram claramente que os descritores utilizados na PCA de alguma forma estão relacionados com a atividade biológica das moléculas. Portanto, investigar os descritores que mais contribuem para PC2 podem fornecer informações úteis para o planejamento de moléculas mais potentes. A fim de alcançar esse objetivo, analisou-se o gráfico de pesos, que mostra a importância relativa de cada descritor para cada PC (Figura 18). Neste gráfico, observa-se que os descritores BAC, BLI, nRCOOH e nROH são mais importantes em PC2 e, por isso, representam características relevantes para a atividade dos compostos.

Por exemplo, o índice central de Balaban (BAC, do inglês *Balaban Centric Index*) é um descritor topológico que descreve a ramificação molecular, ou seja, quanto mais ramificada é a molécula maior o valor do descritor. Levando em consideração que moléculas mais potentes tem valores positivos de PC2, pode-se supor que a ramificação da molécula influencia na complementariedade com seu sítio de ligação. A partir dessa informação, pode-se presumir que moléculas mais ramificadas deveriam ser mais potentes. Entretanto, essa interpretação é incorreta, como discutido mais adiante (vide página 78).

Figura 18. Gráfico de pesos do modelo de PCA demonstrando os descritores mais importantes para PC2.



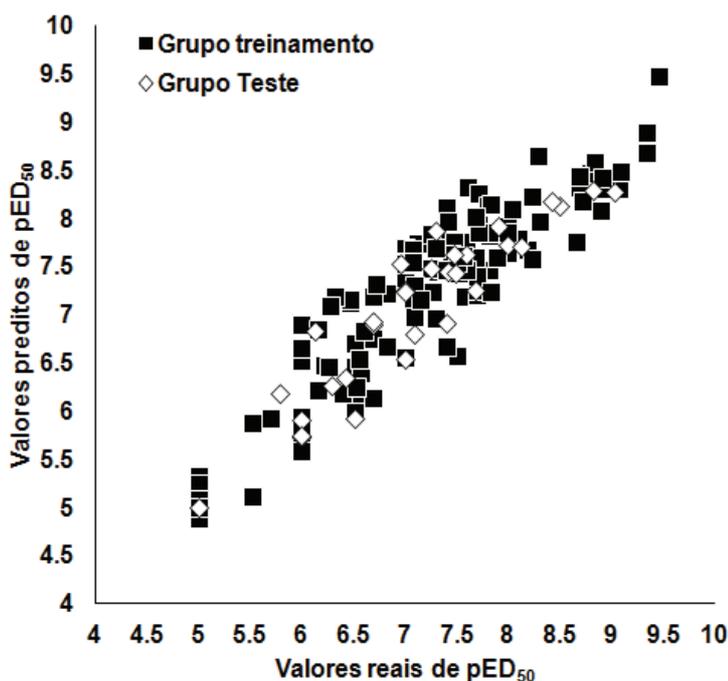
Contudo, a PCA não leva em consideração a atividade biológica para calcular as PCs, portanto características influentes para esse tipo de análise (variância) nem sempre coincidem com propriedades que são importantes para atividade biológica (KOEPPEN et al, 2011). Entretanto, é demonstrado que a utilização de PCs pode classificar o grupo treinamento em subgrupos de moléculas de acordo com a atividade biológica dos compostos e que as variáveis são úteis para tal classificação.

Visando contornar a limitação dos métodos de RLM, foi empregado métodos que utilizam as informações das PCs que selecionam descritores que são ortogonais entre si. Nesse sentido pode-se fazer uma Regressão de Componentes Principais (PCR). Entretanto, a regressão por PLS é uma alternativa mais apropriada por permitir que a informação acerca da atividade biológica (propriedade dependente) seja utilizada no cálculo das PCs, agora denominadas variáveis latentes, visando assim maximizar a covariância entre diferentes conjuntos de variáveis (ROY, ROY, 2008). Dessa forma, pode-se dizer que essa metodologia combina funcionalidades da análise de Componentes Principais e Regressão Linear Múltipla (LEACH, GILLET, 2007), sendo portanto mais robusta que a última (OJHA, ROY, 2011).

A análise PLS é baseada em descritores bidimensionais, que podem ser de difícil interpretação. Entretanto, apesar dessas dificuldades, os modelos de PLS permitem a extração de informações úteis para relação entre a estrutura química com a atividade biológica.

O modelo desenvolvido pela metodologia de PLS, com os 53 descritores oriundos do MobyDigs 1.0, mostrou valores estatísticos bastante satisfatórios ($q^2 = 0,82$; $r^2 = 0,86$) (Figura 19).

Figura 19. Valores reais versus valores preditos do modelo de PLS baseado em descritores bidimensionais com os 53 descritores oriundos do programa MobyDigs 1.0.



Na tentativa de otimizar ainda mais esses parâmetros, descritores com pouca ou nenhuma importância para o modelo foram retirados. Uma ferramenta que auxilia na exclusão de variáveis dos modelos de regressão multivariada é o gráfico de vetor de regressão, que apresenta as variáveis e sua contribuição para o coeficiente de regressão; aquelas com baixos coeficientes de regressão podem ser eliminadas (TEOFILO, MARTINSA, FERREIRA, 2009). Por isso, durante o desenvolvimento do modelo de PLS, descritores que apresentam baixas contribuições para o vetor de regressão foram paulatinamente eliminados, desde que essa estratégia não produzisse diminuição nos valores estatísticos. Com isso, 33 descritores foram descartados. Essa estratégia permitiu a obtenção de um modelo de QSAR com boa consistência interna ($q^2 = 0,86$, $r^2 = 0,9$, com 4 PCs). Embora o melhor modelo obtido apresente valores de q^2 elevados, não é possível afirmar, apenas com base nesse parâmetro, se as informações destacadas nesse modelo são válidas para moléculas congêneras não utilizadas na sua calibração. . Por essa razão, utilizou-se o modelo 5 para prever a atividade das moléculas do conjunto teste, as quais não foram

utilizadas no desenvolvimento do modelo. O poder preditivo encontrado (Equação 1 $r^2_{pred}= 0,84$; Equação 2 $r^2_{pred}= 0,84$) confirma a utilidade do modelo (Tabela 5, Figura 20), sugerindo fornecer informações relevantes sobre propriedades que determinam a atividade biológica desses compostos.

Figura 20. Valores experimentais e preditos de pED_{50} de acordo com o modelo de PLS com 20 descritores bidimensionais.

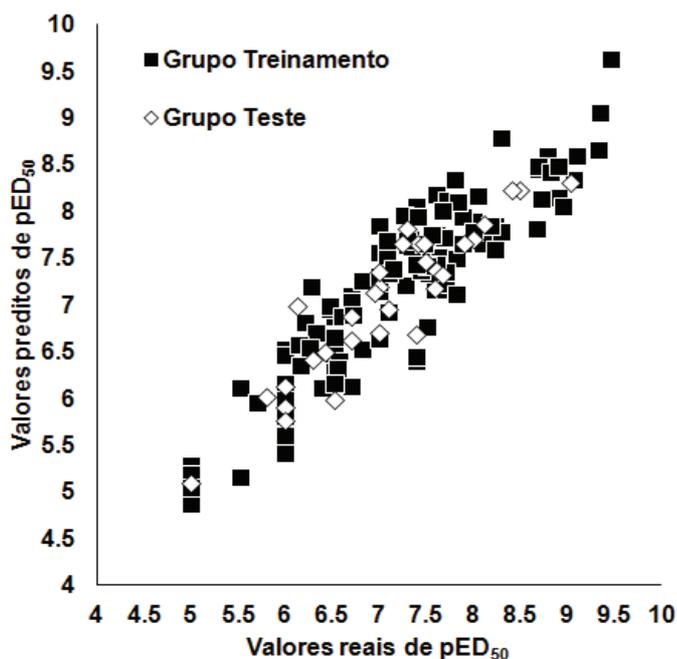


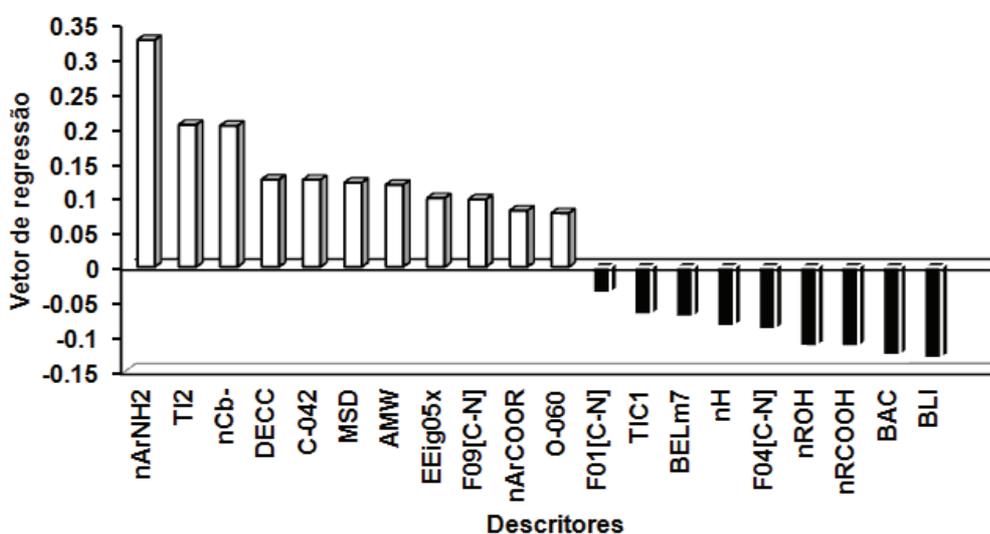
Tabela 5. Valores reais, predito e residuais para os compostos do grupo teste de acordo com o modelo de PLS.

Molécula	Real	Predito	Resíduo	Molécula	Real	Predito	Resíduo
1	6,00	5,90	0,10	17	7,10	6,95	0,15
2	6,52	5,99	-0,47	18	6,42	6,49	-0,07
3	6,12	6,98	-0,86	19	7,60	7,37	0,23
4	7,00	6,70	0,30	20	6,29	6,42	-0,13
5	6,70	6,88	-0,18	21	8,49	8,24	0,25
6	6,00	6,13	-0,13	22	8,12	7,87	0,25
7	5,80	6,01	-0,21	23	7,90	7,66	0,24
8	6,00	5,76	0,24	24	7,68	7,32	0,36
9	5,00	5,10	-0,10	25	9,03	8,31	0,72
10	7,42	7,65	-0,23	26	7,47	7,66	-0,19
11	7,25	7,65	-0,4	27	6,70	6,61	0,09
12	7,00	7,20	-0,20	28	8,82	8,42	0,4

Tabela 5 (continuação). Valores reais, predito e residuais para os compostos do grupo teste de acordo com o modelo de PLS.

13	7,40	6,69	0,71	29	6,96	7,14	-0,18
14	8,00	7,71	0,29	30	7,49	7,47	0,02
15	7,30	7,82	-0,52	31	8,42	8,23	0,19
16	7,00	7,36	-0,36				

A alta qualidade estatística encontrada nos modelos indica que ele pode ser utilizado para a predição da propriedade das moléculas novas. Todavia, além de prever a propriedade de moléculas inéditas, os modelos de QSAR baseados em descritores bidimensionais fornecem informações acerca das propriedades químicas que são relevantes para a atividade dessa classe de moléculas através da análise do gráfico de vetor de regressão (Figura 21 e Tabela 6). Devido ao grande número de descritores presentes no modelo final, apenas aqueles de maior contribuição (positiva e negativa) serão analisados.

Figura 21. Vetor de regressão do modelo de QSAR baseado em descritores bidimensionais.**Tabela 6.** Descritores que contribuem para a atividade dos compostos no modelo de regressão pelo método dos mínimos quadrados parciais.

Descritor	Significado
BAC	Índice central de Balaban
F04[C-N]	Frequência de C-N a uma distância topológica de 04
F01[C-N]	Frequência de C-N a uma distância topológica de 01
NRCOOH	Número de ácidos carboxílicos
NROH	Números de grupos hidroxilas

Tabela 6 (continuação). Descritores que contribuem para a atividade dos compostos no modelo de regressão pelo método dos mínimos quadrados parciais.

BLI	Índice de benzeno de Kier
BELm7	Baixo autovalor n7 da matrix de Burden/peso por massa atômica
TIC1	Índice de informação total (simetria de ordem 1)
Nh	Número de átomos de hidrogênio
nArNH2	Número de aminas primárias ligadas a anel aromático
TI2	Segundo índice de Mohar
MSD	Raiz quadrada da distância do índice de Balaban
nCb ⁻	Número de benzenos substituídos
DECC	Excêntrico
C-042	CH ligado a anel aromático e carbono sp ³
F09[C-N]	Frequência de C-N a uma distância topológica de 09
nArCOOR	Número de ésteres aromáticos
O-060	Al – O – Ar; Ar – O – Ar; R – O – R; R – O – C = X
AMW	Média do peso molecular
EEig05x	Autovalor 05 da matrix de adjacência

O descritor que mais contribui para a atividade é o nArNH₂, um descritor fragmental relacionado ao número de aminas primárias ligada a um anel aromático (TODESCHINI, CONSONNI, 2000). As moléculas mais potentes do conjunto apresentam mais aminas primárias que os de baixa potência (Figura 22, gráfico 1). Entretanto, não é possível afirmar, somente com base nessa informação, se esse efeito é decorrente da capacidade de ligação de hidrogênio, do efeito desse grupo sobre a lipofilicidade das moléculas ou mesmo de uma combinação desses efeitos. O índice de Mohar 2 (TI2, do inglês *second Mohar index*) é um descritor topológico que está relacionado a propriedades estéreas. Quanto menor for o valor desse descritor, menor é a sua potência (Figura 22, gráfico 2).

Entre os descritores que contribuem para a diminuição da atividade está o descritor topológico de índice de aromaticidade (BLI, do inglês *Kier benzene-likeness index*). Ou seja, quanto maior a quantidade de átomos capazes de fazer ressonância na molécula, menor é a potência desse composto (Figura 23, gráfico 3). O descritor BAC, relacionado com a ramificação molecular, também contribui para a diminuição da atividade do grupo de compostos e indica que moléculas mais ramificadas (valores elevados de BAC) tem menor atividade (Figura 23, gráfico 4). Essa informação contradiz as conclusões oriundas da análise de componentes principais e, portanto, mereceu atenção especial de nossa parte.

Figura 22. Importância para o descritor fragmental nArNH2 e TI2 para a atividade biológica de acordo com o modelo de QSAR. Os gráficos 1 e 2 demonstram a relação da variância dos descritores de acordo com pED₅₀ das moléculas. É observável que grupos de moléculas com atividade biológica baixa possuem baixa variância do descritor, enquanto algumas moléculas de com atividade biológica alta contêm alta variância do descritor. Moléculas com alto valor de nArNH2 (gráfico 1 dessa figura, A, nº110 da Tabela 2, com pED₅₀ de 9,45, com duas aminas primárias ligadas ao anel aromático) tem alta atividade biológica, enquanto que a molécula B (nº33 da Tabela 2, com pED₅₀ de 7,61, com 1 amina primária ligada ao anel aromático) tem valor de atividade biológica menor e a molécula C (nº36 da Tabela 2, com pED₅₀ de 6,48, sem aminas primárias) tem baixo valor de atividade. O mesmo raciocínio para o gráfico 2 dessa figura. Moléculas com baixo valor de TI2 (D, nº 83 da Tabela 2) geralmente possuem atividade biológica reduzida, com pED₅₀ de 6,00. Moléculas com alto valor de TI2 (E, nº34 da Tabela 2; F, nº106 da Tabela 2) possuem atividade biológica aumentada, com pED₅₀ de 8,23 e 8,95 respectivamente.

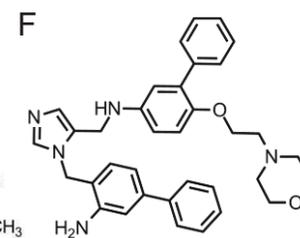
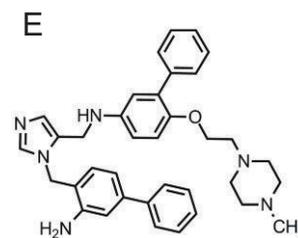
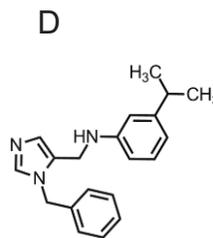
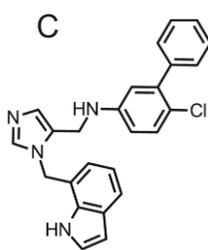
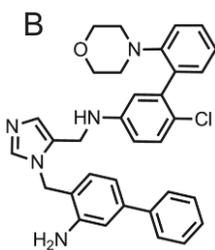
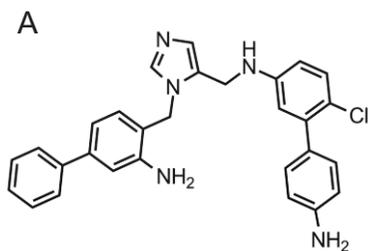
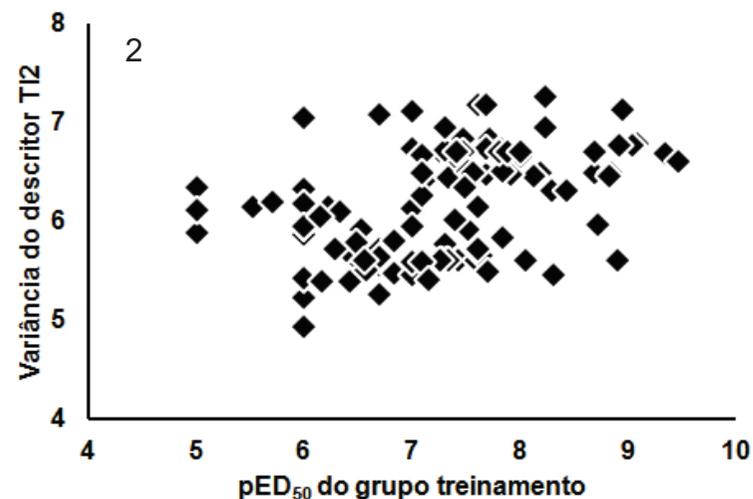
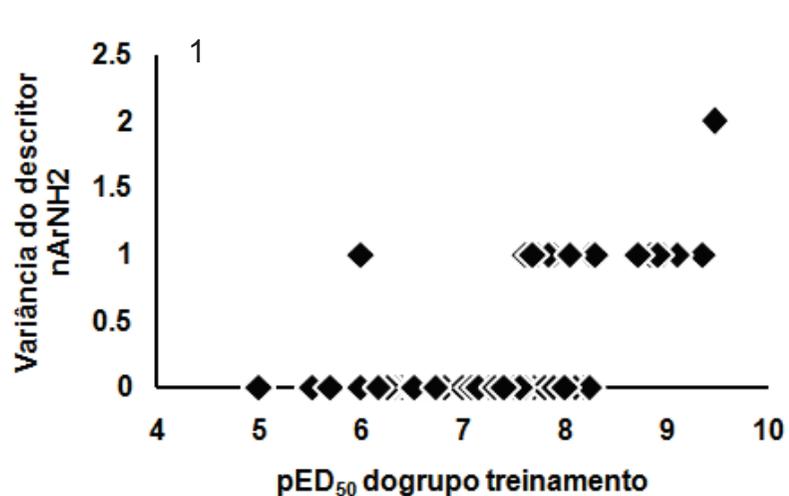
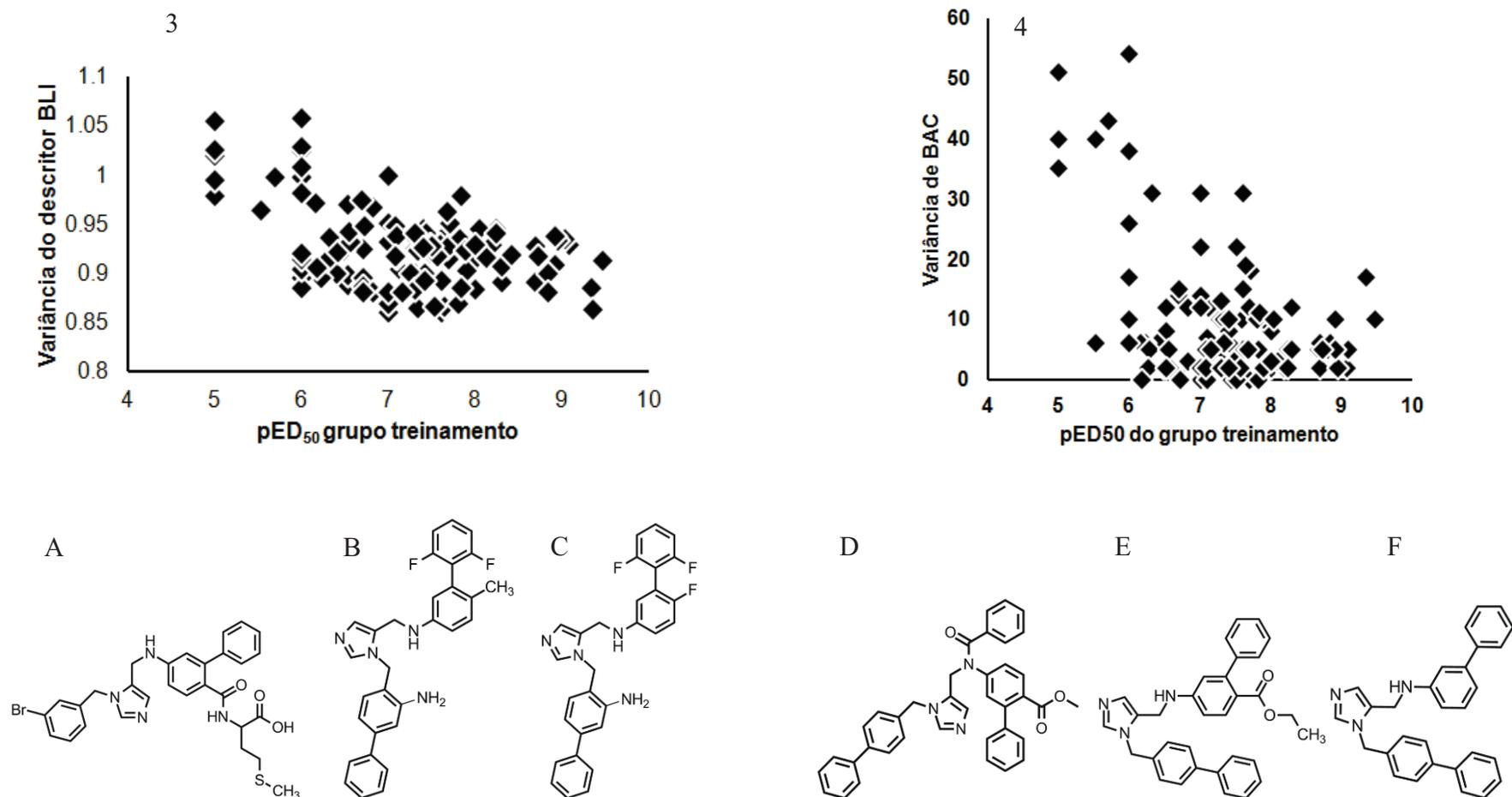
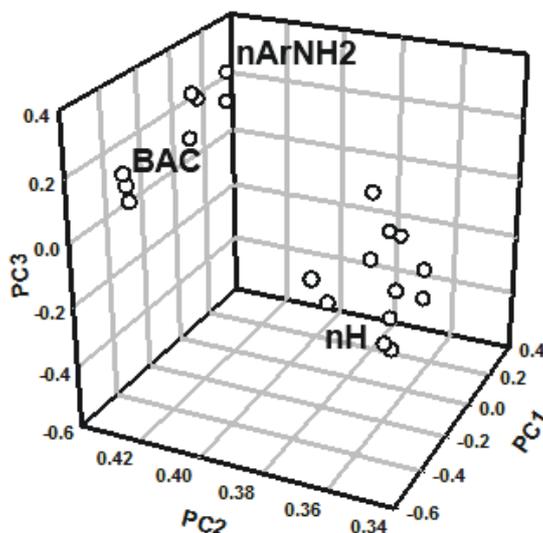


Figura 23. Relação entre a atividade biológica dos inibidores e a variância dos descritores BLI e BAC. Observa-se que moléculas com atividade biológica baixa apresentam elevado valor desse descritor, enquanto que moléculas de atividade biológica alta, esse valor decresce. No gráfico 3 dessa figura a molécula A de nº 31 da Tabela 2 com pED_{50} de 5,00, apresenta várias ligações com átomos passíveis de fazer ressonância e tem valor elevado de BLI. As moléculas B e C, representada pela molécula 108 e 109 da Tabela 2, com pED_{50} 9,34 e 9,35 respectivamente, não possui tantas ligações ressonantes como a molécula A. No gráfico 4 dessa figura, a molécula D está representada uma molécula ramificada (nº 4 da Tabela 2, pED_{50} = 6,52), enquanto que as demais possuem ramificações menores (E nº 61; pED_{50} = 8,00 e F, nº82 da Tabela 2, pED_{50} = 7,69).



A aparente contradição pode ser explicada se considerarmos que o vetor de regressão mostra a contribuição de cada um dos descritores para o modelo final, entretanto ele não mostra a contribuição de cada descritor para as variáveis latentes utilizadas na análise. Na PLS, a orientação das variáveis latentes é diferente daquela encontrada para as Componentes Principais, pois agora o que se quer é maximizar a covariância entre as variáveis independentes (descritores 2D) e a variável dependente (atividade biológica), não mais explicando a variância dos dados originais. Portanto se o gráfico de pesos oriundo da PLS é analisado vemos que os descritores relacionados a PC3 (propriedades estéreas e o descritor fragmental nArNH₂) são importantes para a a atividade (Figura 24).

Figura 24. Importância dos descritores para as PCs segundo modelo de PLS. Considerando PC3, os descritores nArNH₂ e BAC contribuem para a separação dos compostos, enquanto nH tem baixa importância no modelo.



Embora a análise QSAR de descritores topológicos demonstre valores de predição aceitáveis, alguns descritores moleculares são de difícil interpretabilidade (ex. BAC). Portanto, a tentativa de otimização da atividade biológica com base na informação desses descritores pode se tornar uma tarefa complicada. Visando contornar esse tipo de limitação, a técnica de Holograma QSAR (HQSAR) foi empregada.

HQSAR é um método de QSAR simples e robusto que relaciona a atividade biológica com a estrutura molecular, representada em termos de padrão de fragmentos moleculares (LOWIS, 1997; WANG et al., 2003; DA CHEN et al, 2004). Diferentemente da impressão digital, que utiliza seqüências binárias para indicar a

presença ou ausência de um fragmento específico, o holograma molecular mantém a contagem da frequência dos fragmentos. Adicionalmente, ele incorpora informações estereoquímicas não codificadas na impressão digital (GARCIA, HONÓRIO, 2011, MEEK, MOYNA, ZAUHAR, 2008, CHEN et al, 2004, DA CHEN et al, 2004). Essas informações são obtidas a partir da decomposição da estrutura molecular em todos os fragmentos possíveis (linear e ramificados) de M a N átomos (LOWIS, 1997). Esses fragmentos são armazenados num holograma molecular de tamanho pré-definido (53 a 401 posições), que geralmente é menor que o número de fragmentos gerados. Quando isso ocorre é comum que dois fragmentos diferentes, mas semelhantes entre si, ocupem a mesma posição no holograma, gerando assim perda de informação, pois o padrão de ocupação das posições determina as variáveis independentes que serão utilizadas para construção do modelo de QSAR por PLS (TONG et al, 1998; BASKIN, VARNEK, 2008). Outro fator que influencia na alocação dos fragmentos são os parâmetros utilizados para diferenciá-los. Isso pode ser feito levando em consideração sua composição atômica, as ligações entre os átomos, a conectividade (hibridização), os átomos de hidrogênios, a quiralidade e a capacidade dos átomos atuarem como doadores ou aceptores de ligação de hidrogênio. Diante disso, pode-se perceber que modelos de HQSAR são influenciados por três parâmetros: o comprimento do holograma, o tamanho do fragmento e a distinção de fragmentos (AVERY et al, 2002, LOWIS, 1997, SEEL, TURNER, WILLET, 1999).

Em um primeiro momento foi investigada a influência da distinção de fragmento sobre os parâmetros estatísticos dos modelos (Tabela 7). O modelo inicial (A/B/C) apresenta valores estatísticos bons ($q^2=0,76$ e $r^2=0,88$) que são melhorados com a adição dos parâmetro Hidrogênio (H) ($q^2 = 0,80$ e $r^2 =0,90$). Porém, a incorporação do parâmetro quiralidade (Ch) não traz melhoria aos parâmetros estatísticos do modelo (vide modelo 3 da Tabela 7). Por outro lado, considerar simultaneamente a quiralidade e os átomos doadores e/ou aceptores de ligação de hidrogênio (DA) produziu o modelo com melhor ajuste e consistência interna (modelo 4, Figura 23). Esse dado sugere que ligações de hidrogênio são cruciais para atividade biológica

Tabela 7. Influência da distinção de fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com tamanho de fragmento padrão (4 – 7).

Modelos	DF	q^2	r^2	PC	HL
1	ABC	0,76	0,88	6	401
2	ABCH	0,80	0,90	6	307
3	ABCHCh	0,80	0,90	6	307
4	ABCHChDA	0,85	0,91	6	401
5	ABHChDA	0,83	0,91	6	401
6	ABHDA	0,83	0,90	6	401
9	ABCHDA	0,81	0,91	6	307
10	ABCCChDA	0,77	0,88	6	83
11	ABCDA	0,77	0,88	6	307

*Distinção de fragmento: A, átomo; B, ligação; C, conectividade; H, hidrogênio; Ch, quiralidade; DA, doador e aceptor. q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação ; PC, componentes principais; HL, comprimento do holograma.

Posteriormente, foi avaliada a influência do tamanho de fragmento sobre os parâmetros estatísticos do modelo 4 (Tabela 8). Para tal, o tamanho do holograma (401) e distinção de fragmento (A/B/C/H/Ch/DA) foram considerados constantes, variando apenas o valor M e N (mínimo e máximo) do tamanho dos fragmentos (2-5; 3-6; 5-8; 6-9). Observa-se, contudo, que modificações no tamanho padrão não resultaram em melhores valores estatísticos.

Tabela 8. Influência do tamanho dos fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com distinção de fragmento ABCHChDA.

q^2	r^2	PC	HL	TF
0,79	0,88	6	401	2-5
0,84	0,91	6	401	3-6
0,81	0,92	6	401	5-8

* q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação ; PC, componentes principais; HL, comprimento do holograma; TF, tamanho de fragmento.

Para avaliar a estabilidade e significância estatística do modelo, foi realizado o LMO que tem como objetivo analisar de forma mais segura o poder preditivo do modelo retirando um número X de compostos aleatoriamente (Tabela 9) (GRAMATICA, 2007). Se o modelo persistir com valores aproximados de q^2 e r^2

entre os modelos, este é considerado estável. Portanto, está comprovado a estabilidade do modelo.

Tabela 9. Influência do tamanho dos fragmentos sobre os parâmetros estatísticos dos modelos de HQSAR com distinção de fragmento ABCHChDA.

Modelos	LMO	q^2	r^2	PC	HL
1	10	0,85	0,92	6	401
2	15	0,85	0,90	6	401
3	20	0,85	0,91	6	401
4	30	0,85	0,91	6	401

*LMO, Deixar muitos para fora (*Leave-many-out*); q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; PC, componentes principais; HL, comprimento do holograma; TF, tamanho de fragmento.

Para a avaliação da capacidade preditiva do modelo, foi realizado procedimentos de validação externa. O modelo apresentou elevado poder preditivo, tanto na Equação 5 ($r^2_{pred} = 0,82$) quanto na Equação 6 ($r^2_{pred} = 0,82$) (Figura 25, Tabela 10), sugerindo fornecer informações relevantes sobre quais descritores determinam a atividade biológica desses compostos.

Figura 25. Valores reais versus valores preditos de pED_{50} de derivados imidazólicos inibidores da enzima lanosterol 14 alfa-desmetilase de *T. cruzi* no modelo de HQSAR.

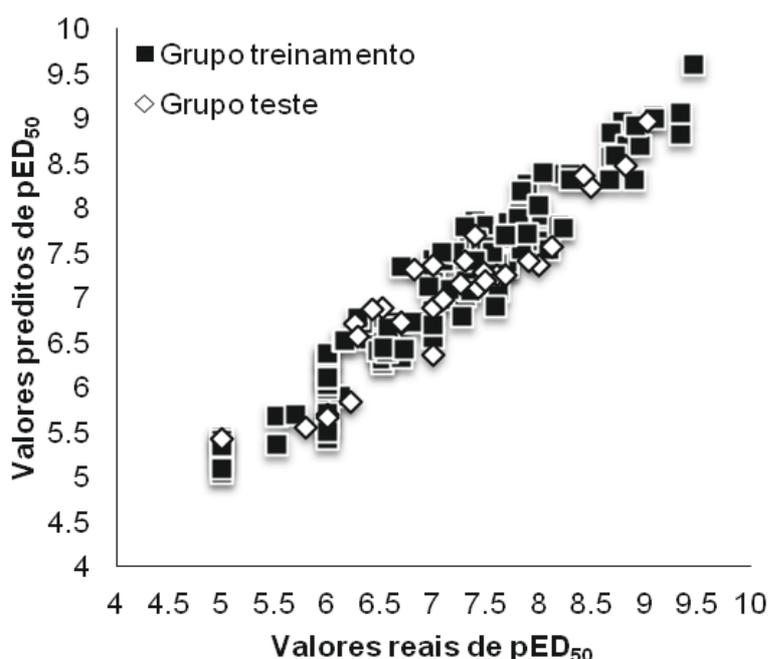


Tabela 10. Valores reais, preditos e residuais para o conjunto teste do modelo A/B/C/H/Ch/DA de tamanho de fragmento 4-7.

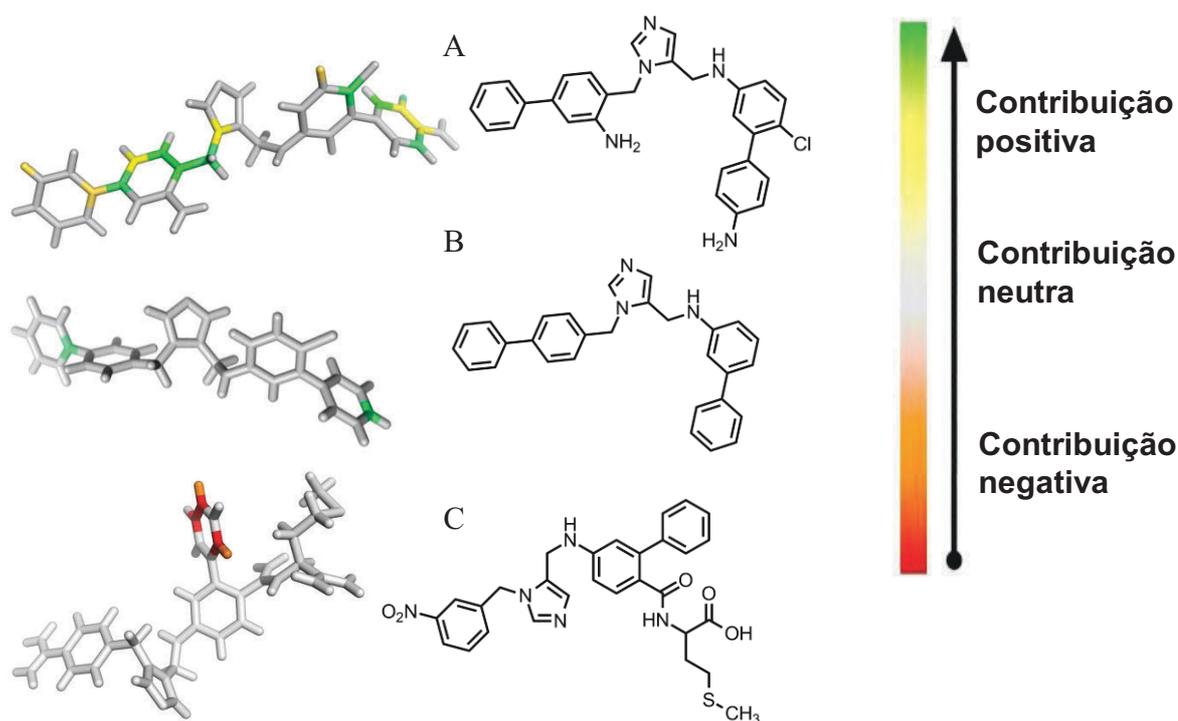
Molécula	Real	Predito	Resíduo	Molécula	Real	Predito	Resíduo
1	6,00	5,68	0,32	17	7,10	7,06	0,03
2	6,52	6,86	-0,34	18	6,42	6,92	-0,5
3	6,12	6,12	0	19	7,60	7,36	0,24
4	7,00	6,32	0,68	20	6,29	6,66	-0,37
5	6,70	6,59	0,11	21	8,49	8,17	0,32
6	6,00	6,75	0,75	22	8,12	7,49	0,63
7	5,80	5,50	-0,30	23	7,90	7,51	0,39
8	6,00	5,60	0,60	24	7,68	7,24	0,44
9	5,00	5,60	-0,60	25	9,03	8,75	0,28
10	7,42	7,07	0,35	26	7,47	7,19	0,28
11	7,25	7,13	0,12	27	6,70	6,58	0,12
12	7,00	7,47	0,47	28	8,82	8,42	0,40
13	7,40	7,75	-0,35	29	6,96	7,42	-0,46
14	8,00	7,46	0,46	30	7,49	7,17	0,32
15	7,30	7,47	-0,17	31	8,42	8,31	0,11
16	7,00	6,97	0,03				

Essa informação pode ser obtida a partir de mapas de contribuição, nos quais cores da extremidade vermelha do espectro (vermelho, vermelho laranja e laranja) são decorrentes contribuições desfavoráveis para a atividade, enquanto que cores da extremidade verde do espectro (amarelo, verde azul e verde) estão relacionadas a contribuições favoráveis. Átomos ou fragmentos com contribuições intermediárias ou que são comuns a todas as moléculas do conjunto estão representados na coloração branca (SALUM, ANDRICOPULO, 2009). Entretanto, eles podem ser importantes para assegurar a orientação correta de grupos farmacofóricos.

A análise dos mapas de contribuição (Figura 26) sugere, por exemplo, que nitrogênios primários ligados a anéis aromáticos contribuem de maneira positiva para a atividade da molécula em relação a moléculas sem esse grupo químico (A vs B). Esse dado está de acordo com dados da literatura que compostos com substituinte amina na posição *orto* do anel melhora a potência antiparasitária (SURYADEVARA et al, 2009). Outro dado observado nos mapas do HQSAR é a

importância da ramificação de determinados grupos. Enquanto que ramificações do anel benzeno na posição *meta* do anel *p*-cloro fenil melhora significativamente a atividade (A), em contrapartida, moléculas com ramificação volumosas, o mesmo anel parece contribuir para redução da atividade biológica (C).

Figura 26. Mapa de contribuição para uma molécula com alta atividade $pED_{50} = 9,45$ (A, nº 110, Tabela 2), uma molécula de atividade intermediária $pED_{50} = 7,69$ (B, nº 82, Tabela 2) e uma molécula de baixa potência $pED_{50} = 5,00$ (C, nº 20, Tabela 2 de acordo com o melhor modelo de HQSAR (A/B/C/H/Ch/DA).

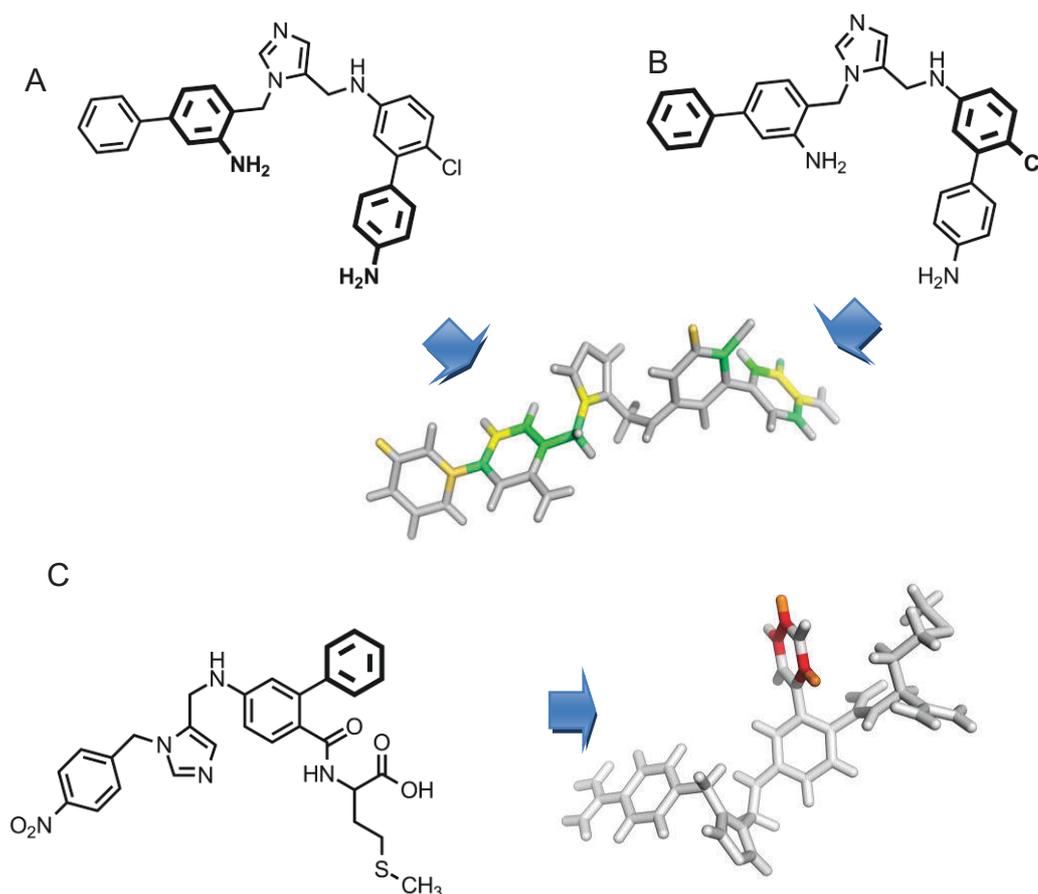


Embora a metodologia HQSAR forneça pistas importantes sobre quais fragmentos estão relacionados com a atividade dos compostos, não é possível determinar diretamente se as contribuições positivas/negativas são decorrentes de fatores eletrônicos ou estereoquímicos. Portanto, visando contornar essa limitação do método, decidiu-se explorar as informações da metodologia de QSAR bidimensional que utilizam descritores topológicos e/ou físico-químicos (MOTA, BARROS, CASTILHO, 2009).

Unificando as duas metodologias, nota-se que determinadas ramificações aromáticas parecem prejudicar a atividade do grupo de compostos, condizentes com o descritor BAC e BLI. Outros grupos químicos podem estar relacionados com aumento da atividade desses compostos. A amina primária da anilina (Figura 27)

está em coerência com o descritor nArNH₂, sendo que os agrupamentos aromáticos pode estar relacionado com a capacidade estérica (TI2).

Figura 27. Interpretação dos modelos de QSAR e QSAR de descritores bidimensionais. Em A está demonstrado em negrito o descritor nArNH₂, correspondente ao mapa de contribuição. Em B, em negrito, pode representar o descritor TI2 e de acordo com parte do mapa de contribuição. Em C, em negrito, pode representar o descritor BAC e BLI, como de acordo com o mapa do modelo de QSAR.



Apesar dos modelos de QSAR 2D se complementarem e permitirem identificar posições relevantes para modificações moleculares, ainda não está claro as razões pelas quais as amins primárias ligadas ao anel aromático contribui positivamente para atividade. Além disso, como a coloração final é resultante da somatória de diversos fragmentos parcialmente sobrepostos, o planejamento de novos derivados somente a partir da análise visual dos mapas de contribuição não é uma tarefa trivial. Adicionalmente, um fator essencial para a atividade dessas moléculas não foi considerado até esse ponto: a complementariedade espacial dos inibidores com seu sítio de interação. Uma alternativa para isso seria construir modelos de QSAR baseados em descritores 3D (MARAN et al, 2010). Todavia, a

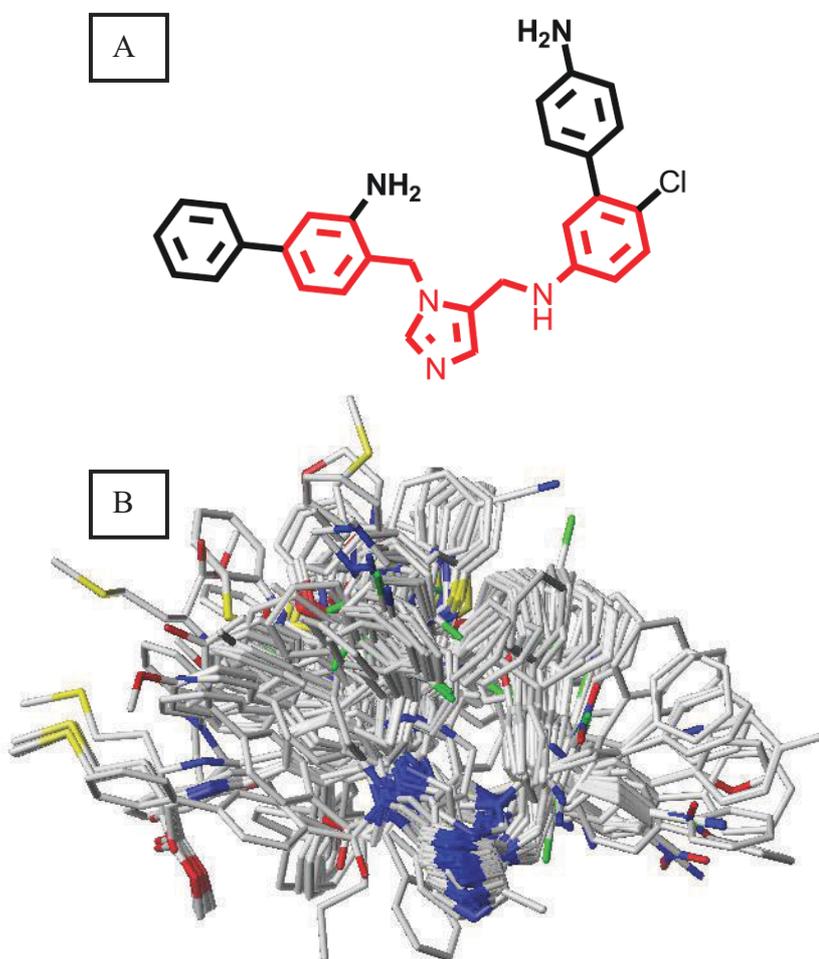
interpretação desse tipo de descritores pode ser tão ou mais complexa que aquela de descritores topológicos, uma vez que além da dificuldade de se encontrar um significado químico para eles, existe ainda a questão de qual conformação deve ser utilizada no cálculo dos descritores. Com o objetivo de resolver, pelo menos parcialmente, esse problema optou-se pela utilização de descritores de campo, que representam a interação de uma carga de prova com as moléculas alinhadas no espaço. Essa metodologia se baseia na possibilidade de descrever as alterações na afinidade de ligação através das interações estéricas e eletrônicas entre os ligantes e a macromolécula. A metodologia que faz uso dessa fundamentação teórica é a Análise Comparativa de Campos Moleculares (CoMFA).

Como mencionado anteriormente, uma etapa fundamental para a construção de um modelo de CoMFA é alinhar as moléculas no espaço de acordo, preferencialmente, com sua conformação bioativa. Na ausência de informações experimentais sobre qual conformação utilizar, é necessário considerar diferentes hipóteses farmacofóricas em busca de um alinhamento molecular que permita desenvolver modelos robustos, preditivos e que sejam úteis na elucidação das relações entre a estrutura química e a atividade biológica. Portanto, os modelos de QSAR 3D (CoMFA) serão apresentados e discutidos conforme a estratégia de alinhamento empregada. Cabe destacar que a fim de permitir comparações e correlações entre os resultados obtidos através de métodos de QSAR 2D e entre os diversos modelos CoMFA, utilizou-se o mesmo conjunto treinamento (Tabela 2) para construção e otimização de todos os modelos de QSAR 3D.

Uma das premissas para a construção de modelos CoMFA é que todas as moléculas devem apresentar o mesmo perfil de interação no sítio ativo da macromolécula. Para tanto, grupos funcionais iguais devem ocupar posições similares no espaço. Por esta razão, a máxima subestrutura comum (MCS) é comumente utilizada no alinhamento estrutural de moléculas. (CUISSART, 2002).

Assim, todas as moléculas do grupo treinamento foram alinhadas na sua conformação minimizada com base nos átomos destacado em vermelho na Figura 28. Durante o alinhamento, as moléculas foram consideradas como rígidas.

Figura 28. Alinhamento dos derivados imidazólicos pelo método da máxima estrutura comum (B). Os grupos químicos considerados comuns a todas as moléculas estão representados em vermelho da molécula A.



As moléculas alinhadas foram, então, inseridas em um cubo que se estende em todas as direções pelo menos 4 Å além das coordenadas das moléculas. Ponteciais de interação esterequímico e eletrônico entre uma carga de prova (C sp³ +1) e os ligantes foram calculadas nas intersecções do cubo (KUBINYI, 1997). Entretanto, quando a carga de prova se sobrepõe aos átomos dos ligantes, as forças repulsivas geradas proporcionam mudanças abruptas nos valores do potencial de Lennard-Jones e de Coulomb. Considerando que em um sistema biológico essa situação não ocorreria, uma vez que a biomacromolécula poderia sofrer uma mudança conformacional para minimizar forças repulsivas entre ela e o ligante, adotou-se um valor máximo para os campos moleculares de interação (30Kcal/mol). Isso evitou que informações irrelevantes fossem utilizadas durante a construção do

modelo (KUBINYI, 1997, KUBINYI, 1998, ZHANG et al, 2011, VERLI, BARREIRO, 2005).

O melhor modelo obtido apresentou bom ajuste ($r^2=0,92$), porém consistência interna insuficiente ($q^2=0,65$). Portanto, na tentativa de produzir modelos com melhores parâmetros estatísticos, recorreu-se a estratégia de “focagem dos campos” (MITTAL, MCKINNON, SORICH, 2009). Na análise padrão do CoMFA todos os pontos da grade molecular tem o mesmo peso na análise de PLS. Entretanto, regiões no espaço onde há grande variação dos valores de campo contribuem mais significativamente para explicar a alteração de atividade nas moléculas em estudo. Por essa razão, cada um dos pontos da grade foi multiplicado pelo coeficiente de desvio padrão naquela região da grade (SDC, do inglês *standard deviation coefficient*). Outro fator que pode influenciar na qualidade estatística do modelo é reduzir o espaçamento dos pontos da grade, pois assim pode-se amostrar com maior precisão as interações da carga de prova com o ligante. A variação desses dois fatores (Tabela 11) resultou numa melhora significativa da consistência interna (modelo 2, com $q^2=0,73$ e $r^2=0,93$). Contudo, como discutido anteriormente, métodos de validação cruzada não são suficientes para garantir que as informações oriundas desse modelo, utilizado para calibração, sejam aplicáveis a moléculas externas. Portanto, a capacidade preditiva do modelo, estimada pelas equação 5 e 6, foi avaliada através da predição da atividade biológica de 31 moléculas não utilizadas na construção do modelo (conjunto teste).

Tabela 11. Resultados do alinhamento de máxima estrutura comum.

Modelos	Parâmetros									
	SDC	Espaçamento	q^2	r^2	SEE	SEP	Teste F	PCs	Fração	
									S	E
1	0,3	0,5	0,69	0,92	0,29	0,59	231,163	6	0,43	0,57
2	0,3	1,0	0,73	0,93	0,28	0,56	246,559	6	0,45	0,55
3	0,3	1,5	0,57	0,72	0,56	0,73	112,982	3	0,48	0,52
4	0,6	0,5	0,69	0,92	0,29	0,59	233,814	6	0,43	0,57
5	0,6	1,0	0,70	0,90	0,33	0,58	183,856	5	0,49	0,51
6	0,6	1,5	0,58	0,72	0,55	0,77	112,111	3	0,50	0,50
7	0,9	0,5	0,68	0,91	0,31	0,61	216,311	6	0,43	0,57
8	0,9	1,0	0,69	0,85	0,41	0,61	140,047	5	0,51	0,49

Tabela 11 (continuação). Resultados do alinhamento de máxima estrutura comum.

9	0,9	1,5	0,57	0,76	0,52	0,69	103,055	4	0,48	0,52
10	1,2	0,5	0,64	0,85	0,41	0,64	137,920	5	0,42	0,58
11	1,2	1,0	0,66	0,81	0,47	0,63	102,535	5	0,50	0,50
12	1,2	1,5	0,55	0,73	0,55	0,71	89,64	4	0,45	0,55

*SDC - coeficiente de desvio padrão; q^2 - coeficiente de correlação da validação cruzada; SEP: erro padrão de predição; r^2 - coeficiente de correlação; *Teste F* - teste de Fisher; *SEE*: erro padrão da estimativa; *PCs* - número de componentes principais; *S* - contribuição estereoquímica; *E* - contribuição eletrostática.

Embora o valor obtido seja razoável ($r^2_{pred} = 0,68$, Equação 5; $r^2_{pred} = 0,68$, Equação 6), se considerarmos a diversidade química presente no conjunto de dados, as conformações utilizadas para construção de modelo não levam em consideração informações acerca do alvo macromolecular, ou seja, esse modelo representa uma super-simplificação da realidade e por essa razão pode levar a conclusões errôneas sobre as exigências estereoquímicas e eletrostáticas das moléculas. Visando averiguar o fundamento dessa hipótese, outras metodologias de alinhamento molecular foram implementadas.

Uma vez que a estrutura tridimensional da biomacromolécula é conhecida, pode ser empregado o método de acoplamento molecular para se encontrar a provável conformação bioativa das moléculas no sítio ativo (KITCHEN, 2004). Atualmente, a maioria dos programas de acoplamento molecular considera a flexibilidade do ligante mantendo a proteína rígida. Há três categorias gerais de algoritmos desenvolvidos para tratar a flexibilidade do ligante: métodos sistemáticos; métodos aleatórios ou estocásticos e métodos de simulação (MAGALHÃES, BARBOSA, DARDENNE, 2007, SOUSA, FERNANDES, RAMOS, 2006). No entanto, como o programa utilizado nesse trabalho (GOLD 4.1.2) utiliza métodos estocásticos, maior enfoque será dado a essa metodologia, em particular a abordagem por algoritmo genético.

No método de algoritmos estocásticos, o espaço conformacional das moléculas é explorado por mudanças aleatórias nos ângulos rotacionais das ligações e posição dos átomos de cada composto, produzindo diversas conformações. Entre as estratégias de busca conformacional que utilizam mecanismos estocásticos pode-se destacar o algoritmo genético que é inspirado no processo biológico da evolução de populações e se baseia no princípio de sobrevivência do indivíduo (conformação) mais adaptado (teoria de Darwin).

Inicialmente, cria-se uma população original, na qual se avalia a aptidão dos indivíduos (moléculas) com posterior reprodução, por cruzamento e mutação, entre os indivíduos mais ajustados a fim de gerar uma nova população (DIAS, AZEVEDO, 2008, ORDOG, GROLMUSZ, 2008, SOUSA, FERNANDES, RAMOS, 2006).

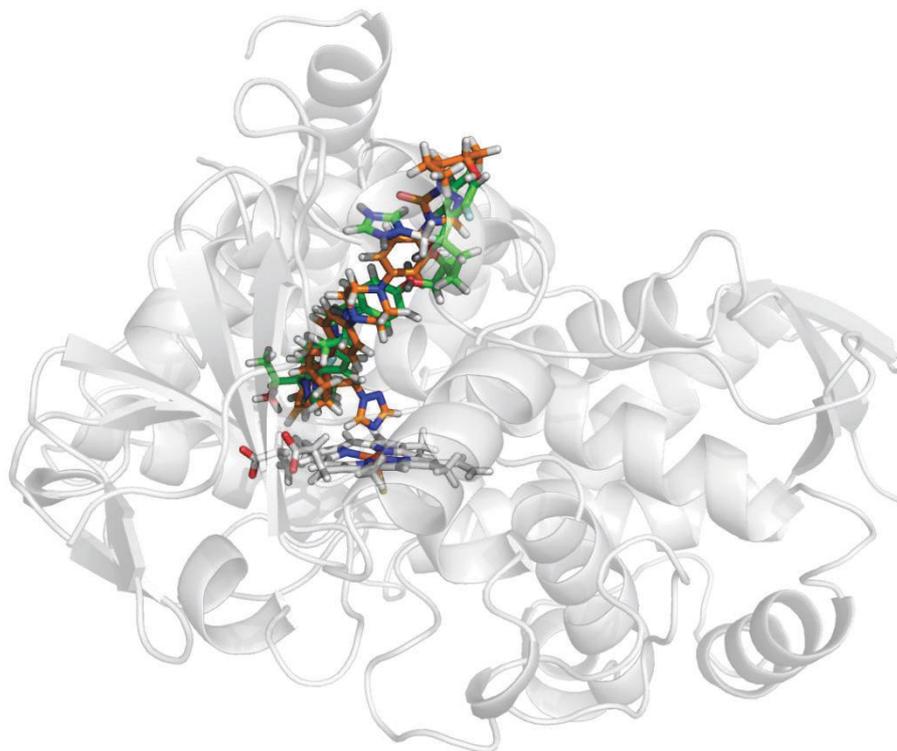
A aptidão das moléculas, mencionada acima é avaliada através de funções de pontuação que devem ser capazes de distinguir os verdadeiros modos de ligação de todos os outros modos gerados através do algoritmo de busca (ABLILASH, 2010). Geralmente as funções de pontuação são classificadas em três classes: baseada no campo de força, empírica e baseada no conhecimento. Cada uma delas tem vantagens e limitações como por exemplo: as funções baseadas em campos de força geralmente não descrevem de forma correta os efeitos de solvatação e termos entrópicos (SOUSA, FERNANDES, RAMOS, 2006, KITCHEN, 2004), enquanto as funções baseadas no conhecimentos, embora descrevam geometrias preferenciais da interação ligante e biomacromolécula, não tem fundamento físico (MAGALHÃES, BARBOSA, DARDENNE, 2007, SOUSA, FERNANDES, RAMOS, 2006). O programa GOLD 4.1.2 apresenta diversas funções de pontuação, como por exemplo, Goldscore e Chemscore (RAHNASTO et al, 2011). A primeira proporciona a avaliação das forças de Van der Waals como termo essencial e usa a estrutura cristalográfica derivada do tratamento de ligação de hidrogênio e termos metálicos. A segunda é uma função de pontuação empírica parametrizada a partir de 82 complexos cristalográficos, cuja afinidade de ligação é conhecida. A função abrange termos de interações lipofílicas, ligações de hidrogênio e penalidade para flexibilidade excessiva (LIEBESCHUETZ, 2012). Devido a dificuldade de se realizar acoplamentos em heme proteínas (formação de complexos de coordenação com o átomo de ferro), foi desenvolvida uma função específica (Goldscore P450) que apresenta maior capacidade de pontuar em primeiro lugar conformações próximas à nativa (KIRTON et al, 2005).

A enzima lanosterol 14 alfa-desmetilase de *T. cruzi* escolhida dentre as que estão depositadas no banco de dados PDB (2WX2 – 2,27 Å; 2WUZ – 2,35 Å; 3KSW – 3,05 Å; 3KHM - 2,85 Å; 3K10 - 2,89 Å) foi a de código 3K10, enzima complexada com o ligante posaconazol. Embora ela tenha alta resolução relacionada as outras, o fator determinante para sua seleção foi o ligante complexado (posaconazol) porque, dentre os ligantes cristalográficos, é o que mais se assemelha aos ligantes desse trabalho em termos de volume. A orientação dos aminoácidos da enzima para o

ligante dentro do sítio ativo e do bolsão proporciona uma análise mais fidedigna, já que estes estão orientados para interação (MAGALHÃES, BARBOSA, DARDENNE, 2007).

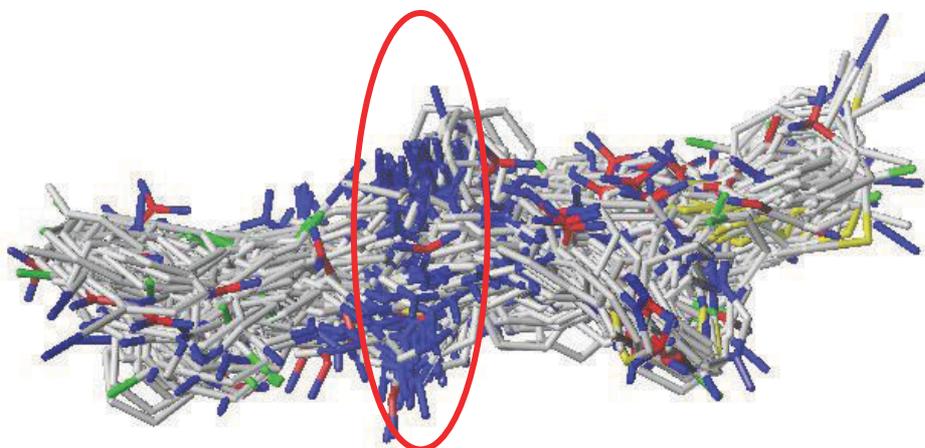
Para testar a qualidade do acoplamento gerado com o auxílio do programa, foi realizado o acoplamento da estrutura cristalográfica do posaconazol nas mesmas condições que o acoplamento do conjunto de dados (Figura 29).

Figura 29. Comparação entre o posaconazol cristalográfico (em laranja) e a solução proposta com o auxílio do programa GOLD 4.1.2 (em verde). Nota-se que a orientação do anel imidazol da solução não interage com o ferro do grupo heme.



Embora o anel orientado em direção ao grupo heme foi o anel triazol trissubstituído que se localiza no final do fármaco, o GOLD 4.1.2 auxiliou na procura de uma solução na qual um anel triazol estivesse próximo ao ferro. Por essa razão, deu-se prosseguimento a metodologia. No entanto, o alinhamento obtido não apresentou, na maioria dos casos, a interação do grupo imidazol com o ferro do grupo heme. As posições adotadas pelos ligantes foram bastante divergentes (Figura 30).

Figura 30. Alinhamento dos compostos do grupo treinamento através do acoplamento molecular sem restrição. Nota-se que alguns anéis imidazóis estão em diferentes locais no espaço (alguns estão inseridos na região circundada em vermelho).



O modelo obtido com parâmetros padrão apresenta ajuste satisfatório ($r^2=0,83$), porém consistência interna muito baixa ($q^2=0,41$). Esse resultado pode ser decorrente de problemas no alinhamento ou então relacionados com o espaçamento da grade e o peso atribuído aos pontos nos quais os campos foram calculados. Visando investigar essa segunda hipótese, a estratégia de focagem foi novamente aplicada (Tabela 12). Embora foi observada uma melhora significativa na consistência interna ($q^2=0,59$), o valor obtido ainda é menor que aquele alcançado para o modelo obtido com moléculas minimizadas (vide página 87). Esse resultado reforça a hipótese de que o principal determinante para os resultados estatísticos insatisfatórios é o alinhamento molecular.

Tabela 12. Valores dos parâmetros estatísticos do alinhamento realizado com o auxílio do programa Gold 4.1.2

Modelos	Parâmetros									
	SDC	Espaçamento	q^2	r^2	SEE	SEP	F test	PCs	Fração	
S									E	
2	0,3	0,5	0,51	0,85	0,40	0,74	175,901	4	0,33	0,67
3	0,3	1,0	0,54	0,78	0,50	0,74	141,052	3	0,36	0,64
4	0,3	1,5	0,29	0,56	0,69	0,89	78,260	2	0,27	0,73
5	0,6	0,5	0,54	0,94	0,25	0,72	332,489	6	0,33	0,67
6	0,6	1,0	0,58	0,84	0,69	0,42	157,835	4	0,39	0,61
7	0,6	1,5	0,26	0,52	0,72	0,90	66,998	2	0,29	0,71

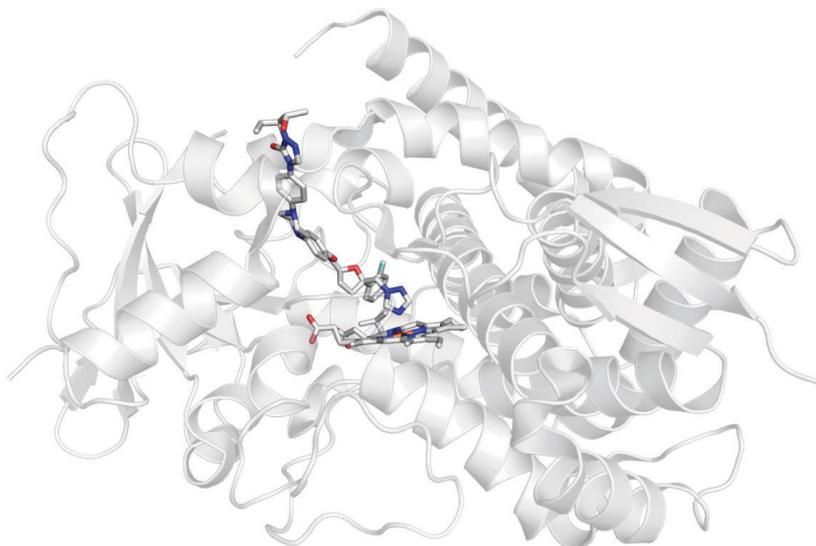
Tabela 12 (continuação). Valores dos parâmetros estatísticos do alinhamento realizado com o auxílio do programa Gold 4.1.2

Modelos	Parâmetros								Fração	
	<i>SDC</i>	<i>Espaçamento</i>	q^2	r^2	<i>SEE</i>	<i>SEP</i>	<i>F test</i>	<i>PCs</i>	<i>S</i>	<i>E</i>
8	0,9	0,5	0,56	0,85	0,41	0,70	167,718	4	0,35	0,65
9	0,9	1,0	0,58	0,83	0,44	0,69	143,649	4	0,40	0,59
10	0,9	1,5	0,26	0,48	0,76	0,92	54,892	2	0,31	0,69
11	1,2	0,5	0,56	0,84	0,42	0,70	158,979	4	0,36	0,64
12	1,2	1,0	0,56	0,84	0,43	0,71	121,976	5	0,39	0,61
13	1,2	1,5	0,27	0,72	0,57	0,91	49,257	6	0,40	0,60

SDC: coeficiente de desvio padrão; q^2 - coeficiente de correlação da validação cruzada; *SEP*: erro padrão de predição; r^2 - coeficiente de correlação; *SEE*: erro padrão da estimativa; *PCs* - número de componentes principais; *S* - contribuição estereoquímica; *E* - contribuição eletrostática.

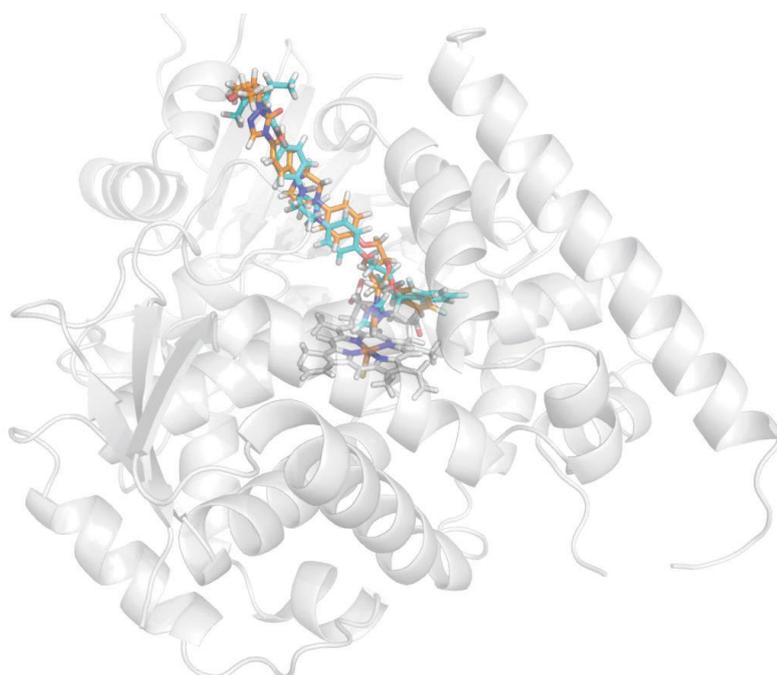
Na tentativa de contornar esse problema, incorporou-se uma restrição na etapa de busca conformacional. De acordo com a estrutura cristalográfica da lanosterol 14 alfa-desmetilase de *T. cruzi* e do ligante posaconazol (código PDB 3K1O) (Figura 31), o complexo apresenta interação do nitrogênio distal do anel imidazol com o átomo de ferro do grupo heme. Considerando que os inibidores em estudo são derivados imidazólicos e que esse anel tem grande propensão a interagir com o átomo de ferro do grupo heme, determinou-se essa restrição. Essa estratégia também é embasada em evidências cinéticas. Foi demonstrado que os derivados se ligam a lanosterol 14 alfa-desmetilase recombinante, produzindo um espectro característico (diferença de absorvância de 430-410) que sugere ligação direta do nitrogênio imidazol em complexo de coordenação com o ferro do grupo heme (SURYADEVARA et al, 2009, BUCKNER, 2003).

Figura 31. Interação entre posaconazol e o ferro do grupo heme da enzima lanosterol 14- α desmetilase de *T. cruzi* (código pdb 3K1O, resolução 2,89).



Para testar a estabilidade da análise foi realizado o reacoplamento do posaconazol no sítio ativo da proteína com a restrição do nitrogênio do anel imidazol com o grupo heme. A estrutura cristalográfica do ligante e a solução proveniente do programa GOLD 4.1.2 são semelhantes (Figura 32), indicando que o método é capaz de buscar soluções plausíveis para os ligantes do grupo total.

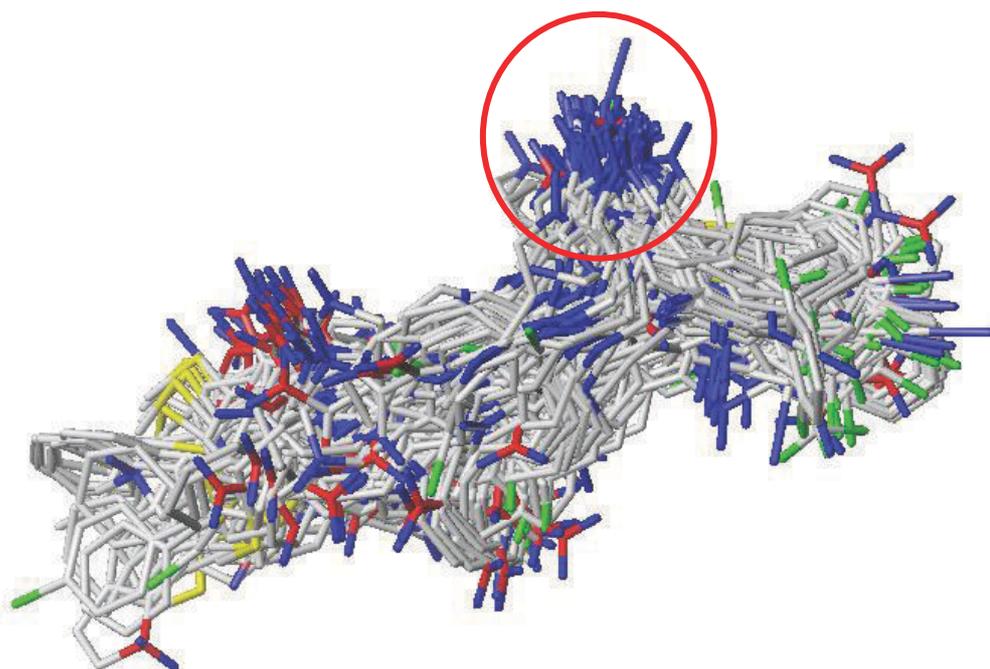
Figura 32. Comparação entre o posaconazol cristalográfico (em laranja) e a solução proposta com o auxílio do programa Gold (em verde). Nota-se que a orientação da solução é semelhante a orientação do ligante cristalográfico.



O novo alinhamento gerado apresentou um padrão característico, sendo que 24,52% das moléculas apresentam interação perpendicular do nitrogênio livre do anel imidazol com o ferro do grupo heme; 29,68% apresentam o nitrogênio do anel imidazol inclinado, porém interagindo com o heme; 27,74% apresentam nitrogênio do anel direcionado ao ferro do grupo heme mas sem interação e 18,01% não apresentam interação. Dessas moléculas que não apresentavam interação, todas possuem o grupamento benzotiazolidínico, cujo nitrogênio faz interação com o ferro, substituindo o imidazol.

Para investigar o modelo de alinhamento baseado na estrutura (Figura 33), foi gerado o modelo de CoMFA. O valor padrão do modelo de CoMFA apresenta valores aceitáveis de ajuste $r^2=0,63$, e consistência interna $q^2= 0,92$.

Figura 33. Alinhamento dos derivados imidazólicos pelo método GOLD. No círculo em vermelho, localiza-se a interação do anel imidazol com o ferro do grupo heme.



Para dar prosseguimento ao processo de otimização do modelo, foi utilizado o procedimento de focagem (Tabela 13). Neste procedimento, o cálculo dos campos de interação pode ser aprimorado com a mudança no espaçamento da grade e/ou com a modificação da importância dos pesos onde os campos estéricos e eletrostáticos (SDC). Os valores estatísticos mostraram-se satisfatórios (modelo 11, com valor de $q^2=0,74$ e $r^2=0,92$ com 5 PCs).

Tabela 13. Resultados do CoMFA obtidos com alinhamento do programa GOLD.

Modelos	Parâmetros								Fração	
	<i>SDC</i>	<i>Espaçamento</i>	q^2	r^2	<i>SEE</i>	<i>SEP</i>	<i>Teste F</i>	<i>PCs</i>	<i>S</i>	<i>E</i>
1	0,3	0,5	0,67	0,94	0,27	0,61	350,200	5	0,33	0,67
2	0,3	1,0	0,72	0,93	0,28	0,55	316,446	5	0,33	0,67
3	0,3	1,5	0,51	0,74	0,54	0,74	113,851	3	0,40	0,60
4	0,6	0,5	0,69	0,92	0,29	0,60	285,581	5	0,33	0,67
5	0,6	1,0	0,74	0,92	0,30	0,54	279,044	5	0,34	0,66
6	0,6	1,5	0,50	0,72	0,56	0,74	100,909	3	0,44	0,56
7	0,9	0,5	0,69	0,92	0,31	0,59	261,763	5	0,33	0,67
8	0,9	1,0	0,72	0,90	0,34	0,56	203,526	5	0,37	0,63
9	0,9	1,5	0,49	0,70	0,59	0,75	89,076	3	0,47	0,53
10	1,2	0,5	0,70	0,92	0,29	0,58	285,019	5	0,30	0,70
11	1,2	1,0	0,72	0,88	0,37	0,57	168,877	5	0,39	0,61
12	1,2	1,5	0,48	0,69	0,59	0,77	65,294	4	0,49	0,51

SDC - coeficiente de desvio padrão; q^2 - coeficiente de correlação da validação cruzada; *SEP*: erro padrão de predição; r^2 - coeficiente de correlação; *SEE*: erro padrão da estimativa; *PCs* - número de componentes principais; *S* - contribuição estereoquímica; *E* - contribuição eletrostática.

Para testar a estabilidade dos modelos de QSAR foi inserido um processo de reamostragem. No processo de reamostragem, a seleção dos grupos de amostras é realizado aleatoriamente e o teste estatístico é calculado usando esses grupos. Os 124 compostos do grupo treinamento foram divididos em 10 grupos gerados aleatoriamente e de tamanho proporcional, sendo executado 10, 20, 50 e 100 vezes durante a análise de PLS. Alguns compostos podem ser incluídos mais de uma vez, enquanto que outros nunca serão selecionados. O parâmetro do modelo é, então, estimado para essa amostra (LEACH, GILLET, 2007). Os resultados da reamostragem demonstram que os modelos gerados não possuem grande variação de r^2_{boot} nem de SEE_{boot} , sendo, portanto, estáveis (Tabela 14).

Tabela 14. Parâmetros estatísticos para análise as estabilidade do modelo do alinhamento baseado na estrutura com 10 grupos escolhidos aleatoriamente.

Número de execuções	Valor médio r^2_{boot}	Valor médio SEE_{boot}
10	0.952 ± 0.011	0.225 ± 0.104
20	0.950 ± 0.009	0.239 ± 0.105
50	0.948± 0.012	0.235± 0.112
100	0.946± 0.011	0.245 ±0.101

r^2_{boot} - $r^2_{bootstrapping}$; SEE_{boot} - erro padrão da estimativa *bootstrapping*.

Além de analisar a preditividade do modelo de validação do grupo treinamento, é importante avaliar a capacidade preditiva dos modelos em relação a compostos que foram completamente excluídos durante sua geração (grupo teste). Portanto, a capacidade preditiva do modelo do grupo treinamento foi avaliada pela previsão de atividades biológicas do conjunto teste. O modelo 5 apresentou $r^2_{pred}=0,70$ (Equação 1); $r^2_{pred}=0,70$ (Equação 2), demonstrando, aparentemente, que é capaz de prever a atividade biológica de suas moléculas e de outras moléculas que não estão sendo exploradas durante a análise (Tabela 15 e Figura 34).

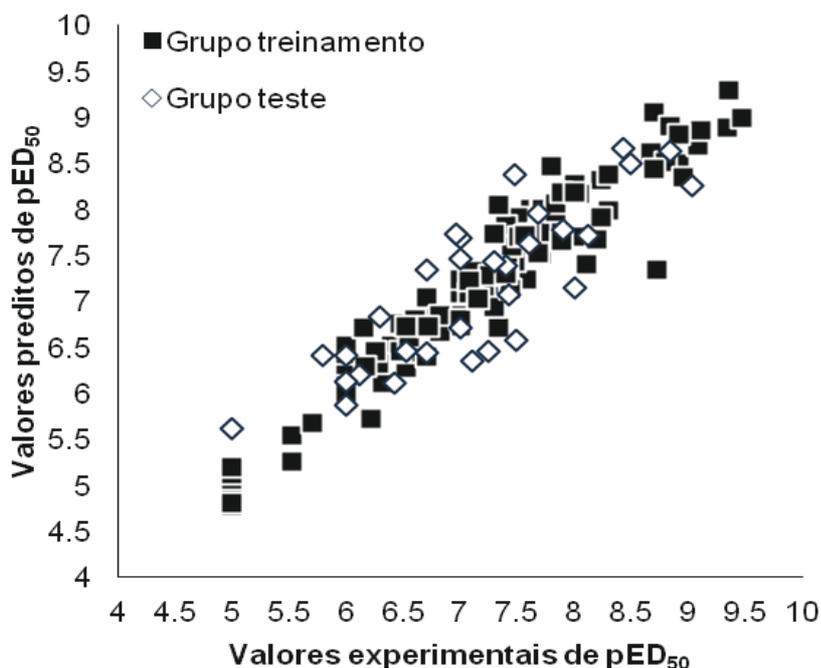
Tabela 15. Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA com o alinhamento do GOLD 4.1.2.

Molécula	Real	Predito	Resíduo	Molécula	Real	Predito	Resíduo
1	6,00	5,87	0,13	17	7,10	6,35	0,75
2	6,52	6,46	0,06	18	6,42	6,11	0,31
3	6,12	6,21	-0,09	19	7,60	7,63	-0,03
4	7,00	7,69	-0,69	20	6,29	6,83	-0,54
5	6,70	6,45	0,25	21	8,49	8,49	0
6	6,00	6,41	-0,41	22	8,12	7,72	0,4
7	5,80	6,41	0,61	23	7,90	7,77	0,13
8	6,00	6,12	-0,12	24	7,68	7,96	-0,28
9	5,00	5,61	-0,61	25	9,03	8,26	0,77
10	7,42	7,07	0,35	26	7,47	8,38	-0,89
11	7,25	6,45	0,8	27	6,70	7,34	-0,64
12	7,00	6,72	-0,28	28	8,82	8,63	0,19
13	7,40	7,39	0,01	29	6,96	7,74	-0,78

Tabela 15 (continuação). Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA com o alinhamento do GOLD 4.1.2

14	8,00	7,14	0,86	30	7,49	6,59	-0,9
15	7,30	7,44	-0,14	31	8,42	8,66	-0,24
16	7,00	7,46	-0,46				

Figura 34. Valores reais vs valores preditos dos inibidores imidazólicos obtidos pelo método de acoplamento molecular.

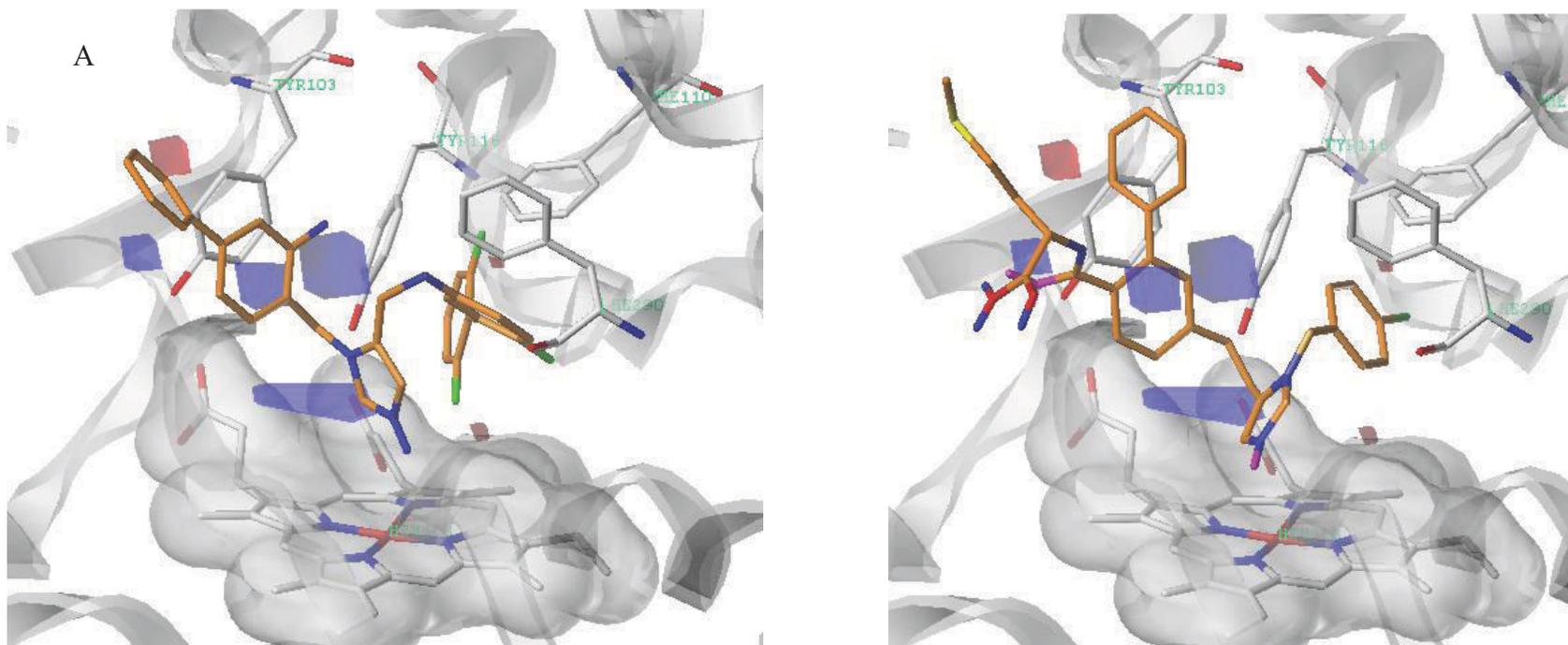


O método de CoMFA permite avaliar visualmente, através dos mapas de contorno, as modificações estruturais relacionadas ao aumento da atividade biológica em estudo. Estes mapas demonstram superfícies circundante aos pontos da grade, onde o modelo de PLS associa mudanças nos valores de interação da carga de prova com a mudança na bioatividade, ou seja, localiza regiões estereoquímicas e eletrostáticas favoráveis e desfavoráveis em determinadas regiões ao redor das moléculas. Os mapas de contornos com coloração azul ou verde exibem regiões que, se exploradas adequadamente, podem aumentar a atividade biológica em decorrência de fatores eletrostáticos ou estereoquímicos, respectivamente. Mapas de contorno que apresentam coloração vermelha ou amarela descrevem regiões onde a carga de prova exibe restrições estereoquímicas e eletrônicas desfavoráveis, respectivamente (MITTAL et al, 2009, MADHAVAN et al, 2012, KUBINYI, 1997). Dessa forma, os mapas de contorno permitem identificar regiões com interações eletrostáticas e/ou estereoquímicas importantes para a

atividade biológica. Por exemplo, o ligante com pED_{50} de 5 (Figura 35-A) apresenta, na posição 19, um anel fenila distal posicionado dentro de um poliedro amarelo. Essa informação sugere que a baixa atividade desse ligante é decorrente do impedimento estereoquímico nessa região. Por outro lado, o ligante com pED_{50} de 9,3 (Figura 35-B) não apresenta esse tipo de impedimento estereoquímico e ainda apresenta um poliedro verde próximo a posição para do anel fenil (posição 6), ou seja, inserir substituintes na posição meta desse anel pode aumentar a potência em decorrência de interações estereoquímicas favoráveis. A análise desses mapas de contorno não indicam se os substituintes devem ter caráter nucleofílico ou eletrofílico. Esse tipo de informação pode ser obtida através dos mapas de contorno eletrostático, que para esse modelo contribuem com 54% da variância total e portanto são mais importantes para explicar a potência dos ligantes. Infelizmente não existem mapas de contorno nessa região do espaço, para o nível de contorno utilizado na análise. Isso pode ser decorrente da ausência ou baixa diversidade química de substituintes nessa posição, a qual resulta em baixa variação nas interações calculadas entre a carga de prova e os ligantes. Portanto, a fim de reduzir essa lacuna de conhecimento seria necessário sintetizar derivados com substituintes nucleofílicos e eletrofílicos nessa posição. Quando se analisa o anel fenila ligado ao nitrogênio do imidazol, nota-se que a amina primária está vizinho a uma região estericamente favorável, próximo a MET^{106} (Figura 36-A), que deve ter caráter deficiente em elétrons, proporcionando essa interação. Essa informação está de acordo com os resultados de HQSAR e QSAR baseado em descritores bidimensionais. Entretanto, a grande vantagem do CoMFA em relação aos modelos bidimensionais está em mostrar que grupos básicos, não apenas a amina primária, devem contribuir positivamente para a atividade.

Embora a estratégia de alinhamento por acoplamento molecular, forçando a interação com o grupo heme, tenha produzido modelos com preditividade apreciável e mapas de contorno que estão de acordo com os resultados de HQSAR, essa metodologia desconsidera as informações referentes a conformação bioativa do ligante cristalográfico, pois ele é removido na etapa de preparação da estrutura para os estudos de acoplamento molecular. Todavia, já foi demonstrado que ensaios virtuais que consideram informações estruturais são mais eficientes que aqueles baseados em acoplamento molecular (HAWKINS, SKILLMAN, NICHOLLS, 2007).

Figura 36. Mapa de contorno do ligante (laranja). Os campos eletrostáticos representado em vermelho e azul devido o cálculo de Coulomb da interação entre a carga de prova e os átomos dos ligantes. Em A está representado um ligante de alta atividade biológica ($pED_{50} = 9,45$, número 110 da Tabela 2); em B ligante de baixa atividade ($pED_{50} = 5,00$, número 20 da Tabela 2).



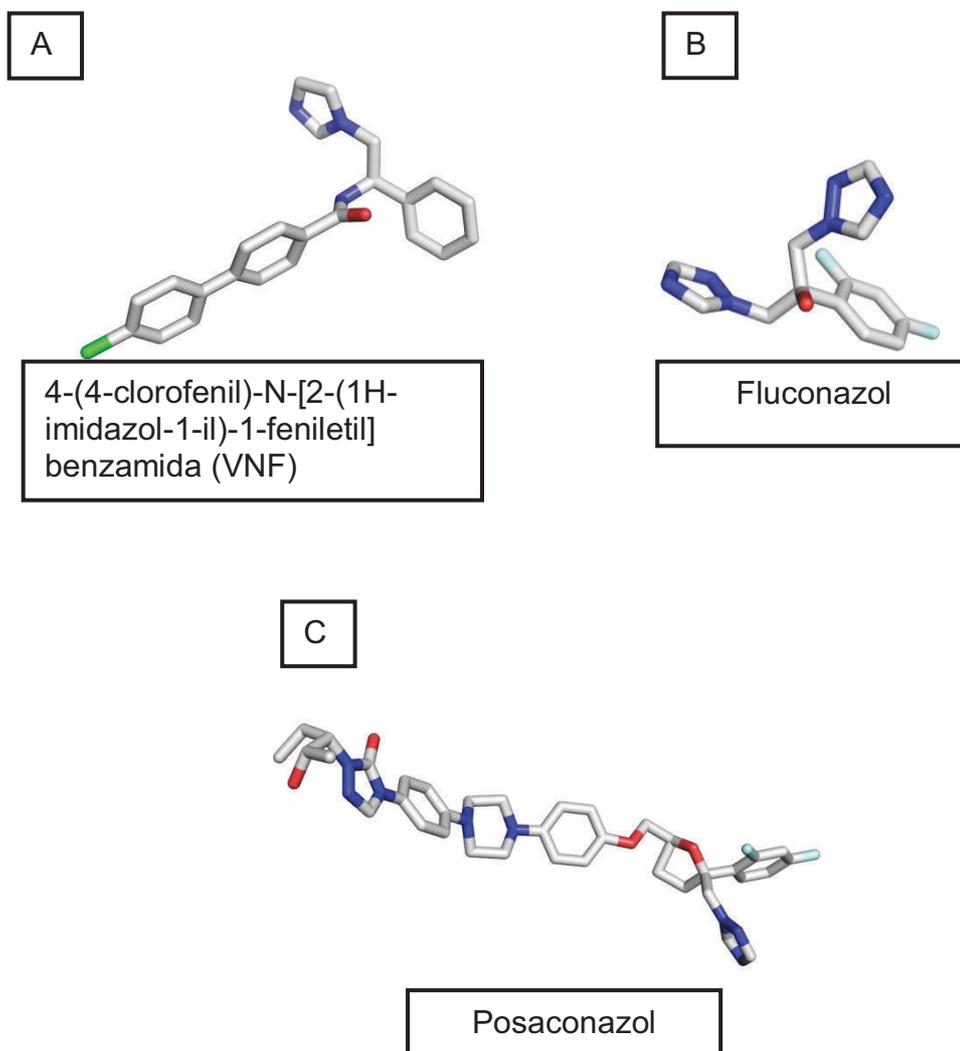
Essa ferramenta é bastante utilizada no planejamento de fármacos. Um exemplo de descoberta de fármaco baseado na estrutura química é a atorvastatina, que foi concebido com base nas informações estruturais do farmacóforo de metabólitos de fungos (lovastatina e mevastatina) e dos hipoglicemiantes semisintéticos (pravastatina e sinvastatina) (WILSON, DANISHEFSKY, 2006, ROTH, 2002).

Seguindo a mesma linha de raciocínio, pode-se supor que alinhamentos moleculares construídos a partir da conformação bioativa do ligante encontrado na estrutura cristalográfica de lanosterol 14 alfa-desmetilase de *T. cruzi* possam dar origem a modelos CoMFA mais robustos e com maior poder preditivo.

Visando investigar essa hipótese foram construídos alinhamentos moleculares através da análise da similaridade morfológica, um método de similaridade molecular 3D que utiliza uma função Gaussiana para calcular a diferença da distância da superfície molecular de dois compostos (JAIN, 2004, CLEVES, JAIN, 2006). Essa informação pode ser utilizada para sobrepor duas ou mais moléculas. Moléculas desalinhadas mas com algum grau de semelhança entre si apresentam característica semelhantes na sua superfície, se analisada a partir de pontos de observação, distribuídos ao redor da cada molécula. Conjunto de pontos externos (triângulo) podem ser utilizados como uma representação simplificada da molécula e nesse caso o alinhamento molecular limita-se a encontrar a melhor sobreposição dos triângulos (Figura 2) (JAIN, 2004, JAIN, 2000).

A forma mais simples de implementar essa estratégia de alinhamento é através da comparação morfológica das moléculas do conjunto de dados e um ligante da lanosterol 14 alfa-desmetilase de *T. cruzi* que tenha sido cristalizado em sua conformação bioativa no sítio ativo da enzima. Entre as alternativas disponíveis (VNF, encontrado na estrutura 3KSW; fluconazol, encontrado na estrutura 2WX2, 2WUZ e 3KHM e posaconazol, encontrado na estrutura 3K1O) (Figura 37), optou-se pelo posaconazol. Essa escolha foi guiada principalmente pelo fato desse fármaco apresentar volume molecular semelhante ao do grupo de compostos deste trabalho.

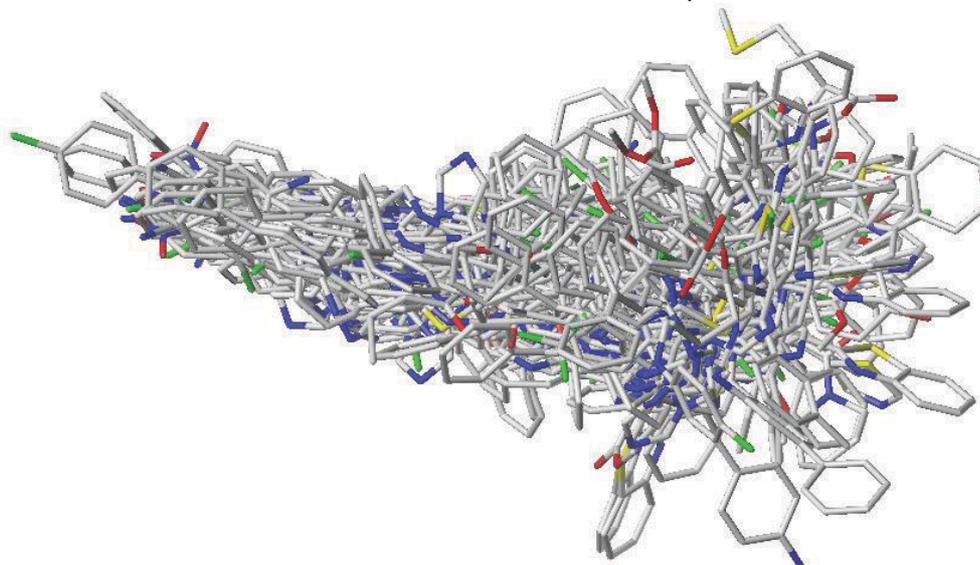
Figura 37. Ligantes cristalizados que interagem com a enzima Lanosterol 14 alfa-desmetilase de *T. cruzi*. Em A, o ligante 4-(4-clorofenil)-N-[2-(1H-imidazol-1-il)-1-feniletil] benzamida (VNF), que interage com uma parte específica do sítio no qual nenhum outro ligante interage (hélice B'/loop B'C; hélice C e porção N terminal da hélice I). Em B, a molécula do fluconazol e em C a do posaconazol.



Fonte: Adaptação de CHEN et al, 2010, LEPESHEVA, 2010.

A sobreposição (similaridade morfológica) das moléculas (Figura 38) pode ser avaliada numericamente. A pontuação no valor de 10,0 (dez) é atribuída para moléculas que se sobrepõe perfeitamente a molécula referência (moléculas idênticas) enquanto que o valor 0,0 (zero) refere-se a moléculas totalmente diferentes em relação a forma e polaridade (CLEVES, JAIN, 2008). O conjunto de moléculas desse trabalho foi sobreposto ao posaconazol pela metodologia de similaridade morfológica. As moléculas do conjunto treinamento apresentam valores de 3,92 (ligante número 123 da Tabela 2) a 5,5 (ligante 90 da Tabela 2).

Figura 38. Alinhamento dos derivados imidazólicos sobre o posaconazol.



O modelo CoMFA gerado a partir desse alinhamento apresentou parâmetros estatísticos inferiores ($q^2 = 0,35$ e $r^2 = 0,91$ com 4PCs) aos do modelo obtido a partir do acoplamento molecular, mesmo após a estratégia de focagem ter sido aplicada (modelo 8 com $q^2 = 0,60$; $r^2 = 0,92$ com 5PCs Tabela 16).

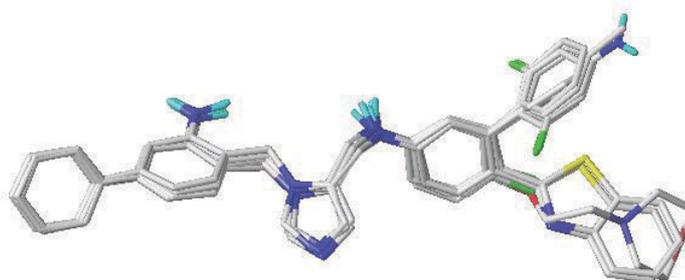
Tabela 16. Resultados do CoMFA com alinhamento do Posaconazol pela metodologia de similaridade morfológica realizado com o auxílio do programa Surfex-sim.

Modelos	Parâmetros								Frações	
	SDC	Espaçamento	q^2	r^2	SEE	SEP	F test	PCs	S	E
1	0,3	0,5	0,43	0,92	0,30	0,80	328,537	4	0,34	0,66
2	0,3	1,0	0,56	0,91	0,33	0,70	285,542	4	0,37	0,63
3	0,3	1,5	0,29	0,81	0,46	0,89	127,792	4	0,42	0,53
4	0,6	0,5	0,46	0,94	0,26	0,78	376,351	5	0,34	0,66
5	0,6	1,0	0,60	0,92	0,29	0,68	283,487	5	0,42	0,58
6	0,6	1,5	0,29	0,54	0,71	0,88	70,630	2	0,54	0,46
7	0,9	0,5	0,46	0,89	0,35	0,78	238,336	4	0,35	0,65
8	0,9	1,0	0,60	0,93	0,28	0,68	257,107	6	0,46	0,54
9	0,9	1,5	0,29	0,47	0,77	0,88	52,716	2	0,64	0,36
10	1,2	0,5	0,46	0,83	0,44	0,78	190,222	3	0,38	0,62
11	1,2	1,0	0,58	0,88	0,37	0,69	172,013	5	0,49	0,51
12	1,2	1,5	0,29	0,43	0,79	0,89	45,664	2	0,69	0,31

SDC - coeficiente de desvio padrão; q^2 - coeficiente de correlação da validação cruzada; SEP: erro padrão de predição; r^2 - coeficiente de correlação; SEE: erro padrão da estimativa; PCs - número de componentes principais; S - contribuição estereoquímica; E - contribuição eletrostática.

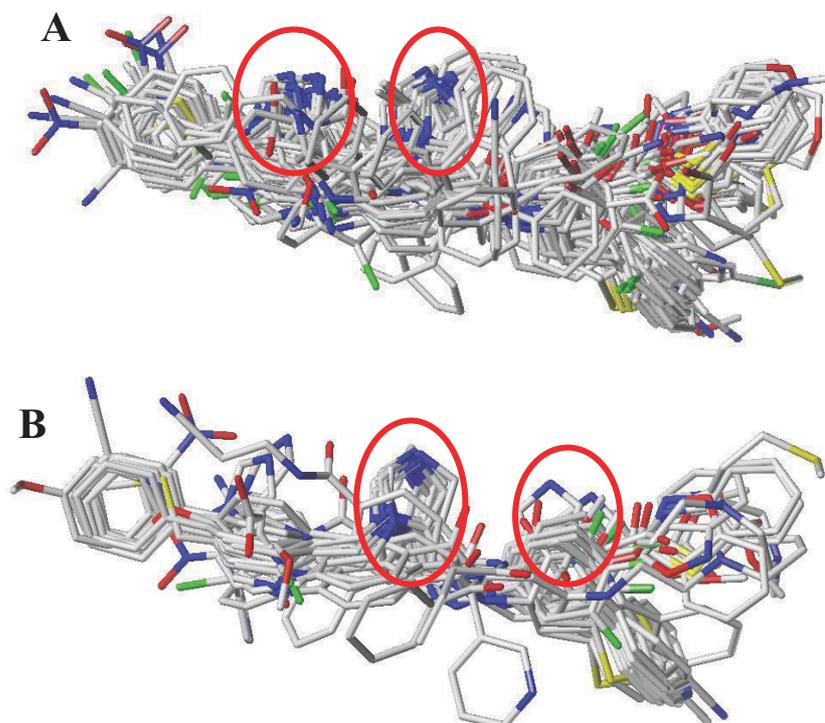
Um dos fatores que pode estar influenciando negativamente os modelos é a baixa similaridade química entre as moléculas do conjunto treinamento e a molécula referência, já que a similaridade média dos compostos é de 46% (pontuação média 4,64). Visando contornar esse problema, uma nova estratégia de alinhamento foi implementada: considerando que os derivados imidazólicos em estudo devem adotar conformação similar quando no sítio ativo da lanosterol 14 alfa-desmetilase, foi realizado um alinhamento múltiplo dos 5 inibidores de maior atividade biológica do grupo treinamento com base na sua similaridade morfológica. Essa estratégia tem como vantagem a formação de um grupo farmacofórico de diversos ligantes de alta atividade biológica com suas principais características. Durante esse processo as moléculas foram consideradas flexíveis e não houve restrição de qual molécula deveria ser utilizada como molde. A sobreposição molecular de maior pontuação (Figura 39) apresenta os grupos químicos bem sobrepostos em termos de volume e características atômicas.

Figura 39. Modelo de similaridade das 5 moléculas mais potentes do conjunto de dados.



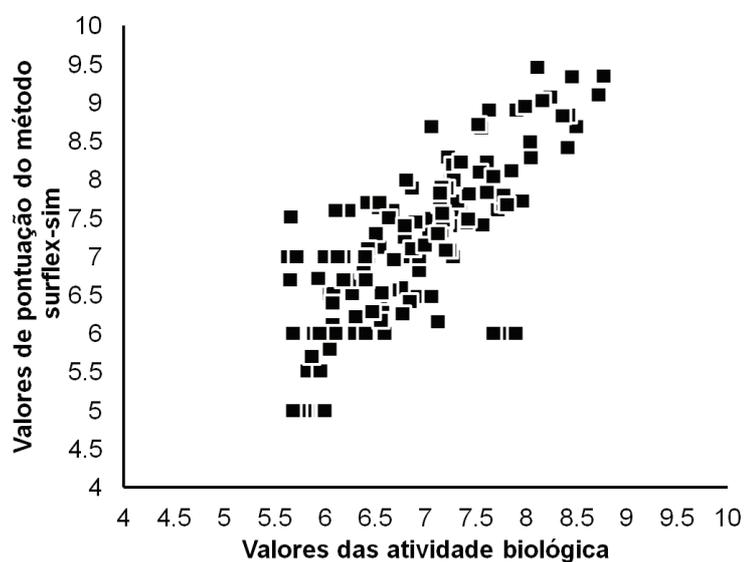
Em um segundo momento, a hipótese gerada das 5 moléculas de maior atividade biológica foi utilizada como molde para a sobreposição molecular dos demais compostos deste trabalho e posterior alinhamento para o CoMFA. Neste alinhamento observa-se que os anéis imidazóis concentram-se em duas regiões no espaço (Figura 40). Esse resultado pôs em dúvida a hipótese do alinhamento por similaridade molecular, pois se sabe que os derivados bifenílicos imidazóis interagem com o ferro do grupo heme e que provavelmente estariam localizados em uma mesma região do espaço.

Figura 40. Alinhamento dos derivados imidazólicos pelo método surflex-sim. Representação em A do conjunto treinamento e em B do conjunto teste. Em círculo vermelho, a localização dos anéis imidazol.



Essa aparente contradição pode ser explicada quando se analisa os valores de pontuação de sobreposição das conformações versus a atividade biológica dos derivados imidazólicos. Observa-se claramente que há uma correlação entre esses valores (Figura 41) e que as moléculas que apresentam o anel imidazol na posição alternativa são as de menor potência.

Figura 41. Relação das pontuações das conformações geradas com o auxílio do programa Surfex-sim versus pED_{50} dos compostos.



O modelo CoMFA construído a partir desse alinhamento apresenta ótimo ajuste ($r^2=0,92$) mas consistência interna limitada ($q^2=0,60$), com 6 PCs. Na tentativa de otimização do modelo, aplicou-se novamente a estratégia de focagem do modelo. Essa estratégia resultou no aumento da consistência interna ($q^2= 0,75$) e manutenção do ajuste do modelo ($r^2= 0,91$) (Tabela 17, modelo 5).

Tabela 17. Resultados estatísticos dos modelos CoMFA obtidos a partir do alinhamento por similaridade química utilizando os 5 inibidores mais potentes.

Modelos	Parâmetros								Fração	
	SDC	Espaçamento	q^2	r^2	SEE	SEP	F test	PCs	S	E
1	0,3	0,5	0,66	0,93	0,28	0,62	271,815	6	0,44	0,56
2	0,3	1,0	0,73	0,92	0,30	0,55	239,377	6	0,49	0,51
3	0,3	1,5	0,55	0,67	0,60	0,70	121,876	2	0,40	0,60
4	0,6	0,5	0,71	0,93	0,28	0,58	266,003	6	0,46	0,54
5	0,6	1,0	0,75	0,91	0,31	0,53	208,204	6	0,52	0,48
6	0,6	1,5	0,55	0,66	0,61	0,71	117,294	2	0,43	0,57
7	0,9	0,5	0,72	0,93	0,29	0,56	246,601	6	0,47	0,53
8	0,9	1,0	0,74	0,90	0,34	0,54	172,943	6	0,53	0,47
9	0,9	1,5	0,53	0,65	0,62	0,72	110,335	2	0,46	0,54
10	1,2	0,5	0,73	0,92	0,30	0,56	220,208	6	0,48	0,52
11	1,2	1,0	0,73	0,87	0,38	0,55	156,749	5	0,55	0,45
12	1,2	1,5	0,52	0,63	0,64	0,73	102,710	2	0,49	0,51

SDC - coeficiente de desvio padrão; q^2 - coeficiente de correlação da validação cruzada; SEP: erro padrão de predição; r^2 – coeficiente de correlação; SEE: erro padrão da estimativa; PCs – número de componentes principais; S - contribuição estereoquímica; E – contribuição eletrostática.

Para avaliar a estabilidade do modelo, a função de reamostragem (*bootstrapping*) foi usada. A determinação do erro nos valores de r^2 (r^2_{boot}) e o cálculo do erro padrão (SEE_{boot}) de cada modelo (Tabela 18), juntamente com a determinação do valor do Teste F (208,204) confirmam a estabilidade e significância estatística do modelo (GUIDO et al, 2008).

Tabela 18. Valores de r^2_{boot} e SEE_{boot} para avaliar a estabilidade do modelo de similaridade morfológica do modelo dos cinco ligantes mais potentes.

Número de execuções	r^2_{boot}	SEE_{boot}
10	0.945±0,012	0,249±0,102
20	0.946±0.012	0,242 ±0,100
50	0.945± 0.009	0.247 ±0.094
100	0.945 ± 0.010	0.248 ±0.092

r^2_{boot} - $r^2_{bootstrapping}$; SEE_{boot} - erro padrão da estimativa bootstrapping.

No entanto, como foi discutido anteriormente, avaliar apenas a consistência interna do modelo por validação cruzada é insuficiente para assegurar a aplicabilidade dos mesmos no planejamento de inibidores mais potentes. Por essa razão, a habilidade preditiva da atividade biológica dos modelos foi avaliada para as moléculas do grupo teste (Tabela 3), que foi completamente excluído da calibração do modelo. Os valores de r^2_{pred} encontrados (0,79 - Equação 5; 0,79 - Equação 6) comprovam que esse modelo é superior aos demais no que se refere a capacidade preditiva (Tabela 19, Figura 42). Diante desse resultado, decidiu-se analisar os mapas de contorno fornecidos por esse modelo e compará-lo com aqueles gerados através do alinhamento por acoplamento molecular, no qual a interação com o grupo heme foi imposta.

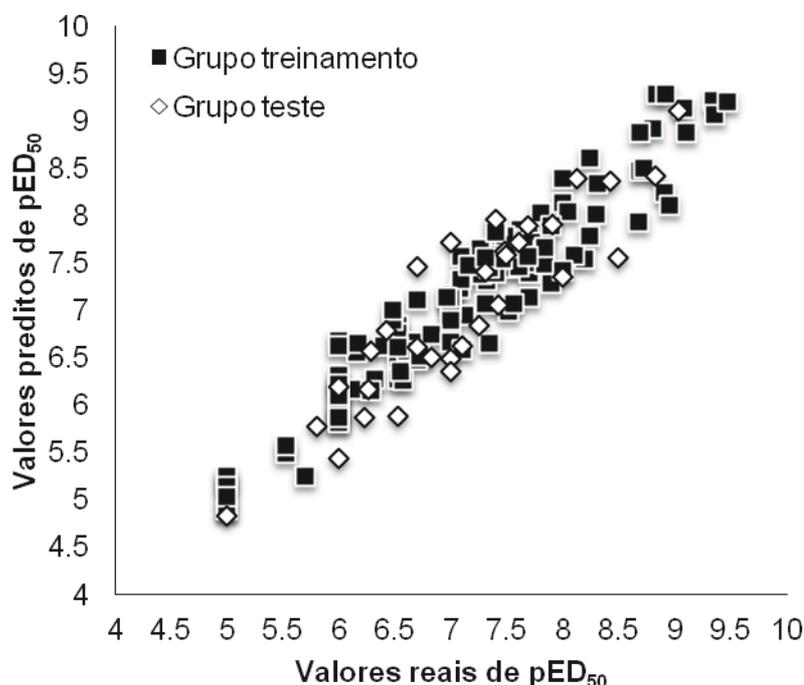
Tabela 19. Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA com o alinhamento do Surflex-sim.

Molécula	Real	Predito	Resíduo	Molécula	Real	Predito	Resíduo
1	6,00	5,99	0,01	17	7,10	6,67	0,43
2	6,52	5,99	0,53	18	6,42	6,90	-0,48
3	6,12	6,92	-0,8	19	7,60	7,65	-0,05
4	7,00	6,57	0,43	20	6,29	6,64	-0,35
5	6,70	7,49	-0,79	21	8,49	7,60	0,89
6	6,00	6,04	-0,04	22	8,12	8,38	-0,26
7	5,80	5,92	-0,12	23	7,90	7,92	-0,02
8	6,00	6,25	-0,25	24	7,68	7,95	-0,27
9	5,00	4,82	0,18	25	9,03	9,07	-0,04
10	7,42	7,04	0,38	26	7,47	7,59	-0,12

Tabela 19 (continuação). Valores reais, preditos e residuais para o conjunto teste do modelo CoMFA com o alinhamento do Surfex-sim.

11	7,25	6,80	0,45	27	6,70	6,79	-0,09
12	7,00	7,75	-0,75	28	8,82	8,42	0,40
13	7,40	8,03	-0,63	29	6,96	7,42	-0,46
14	8,00	7,38	0,62	30	7,49	7,60	-0,11
15	7,30	7,47	-0,17	31	8,42	8,34	0,08
16	7,00	6,37	0,63				

Figura 42. Gráfico real versus predito do modelo de CoMFA com alinhamento do Surfex-sim.



O mapa de contorno estereoquímico (Figura 43), que explica 52% da variância total dos dados (Tabela 17), sendo portanto mais importante para atividade, sugere de forma mais evidente que a amina primária tem contribuição estérea positiva, corroborando com as informações do HQSAR, QSAR baseado em descritores bidimensionais assim como o modelo de CoMFA com alinhamento baseado na biomacromolécula.

Substituições na posição meta do anel *p*-cloro fenil também aumenta a atividade (Figura 43-A). Por outro lado, verifica-se que não pode ocorrer substituição no anel imidazol da molécula de maior atividade devido a uma região estericamente

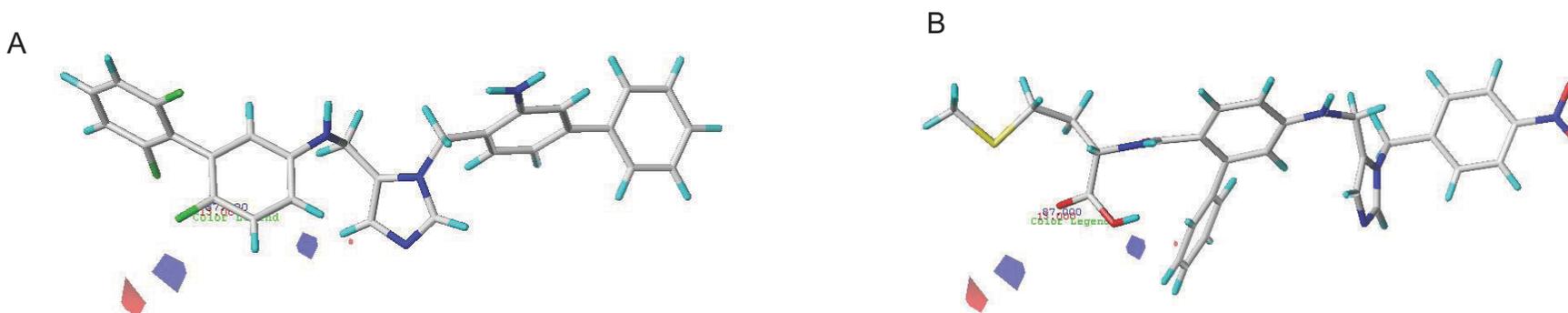
desfavorecida e que moléculas com baixa atividade biológica (Figura 43-B) tem o anel imidazol posicionado nesta região.

A análise dos mapas de contorno eletrostático, por sua vez, indica que substituintes nucleofílicos devem ocupar a posição meta do anel *p*-cloro fenil (Figura 44). Essa informação também auxilia a entender por que a molécula 20 (Tabela 2) apresenta baixa atividade biológica, uma vez que ela apresenta um carboxilato próximo de uma região onde grupos ricos em elétrons não são favoráveis.

Figura 43. Mapa de contorno estereoquímico dos derivados do modelo de CoMFA do alinhamento do Surfex-sim. (A) Molécula com alta atividade biológica, com pED_{50} de 9,45; (B) Molécula com baixa atividade, com pED_{50} de 5,00.



Figura 44. Mapa de contorno eletrostático dos derivados do modelo de CoMFA do alinhamento do Surfex-sim. (A) Molécula com atividade alta, com pED_{50} de 9,45 e (B) Molécula com baixa atividade, com pED_{50} de 5,00.



5.0 CONCLUSÃO

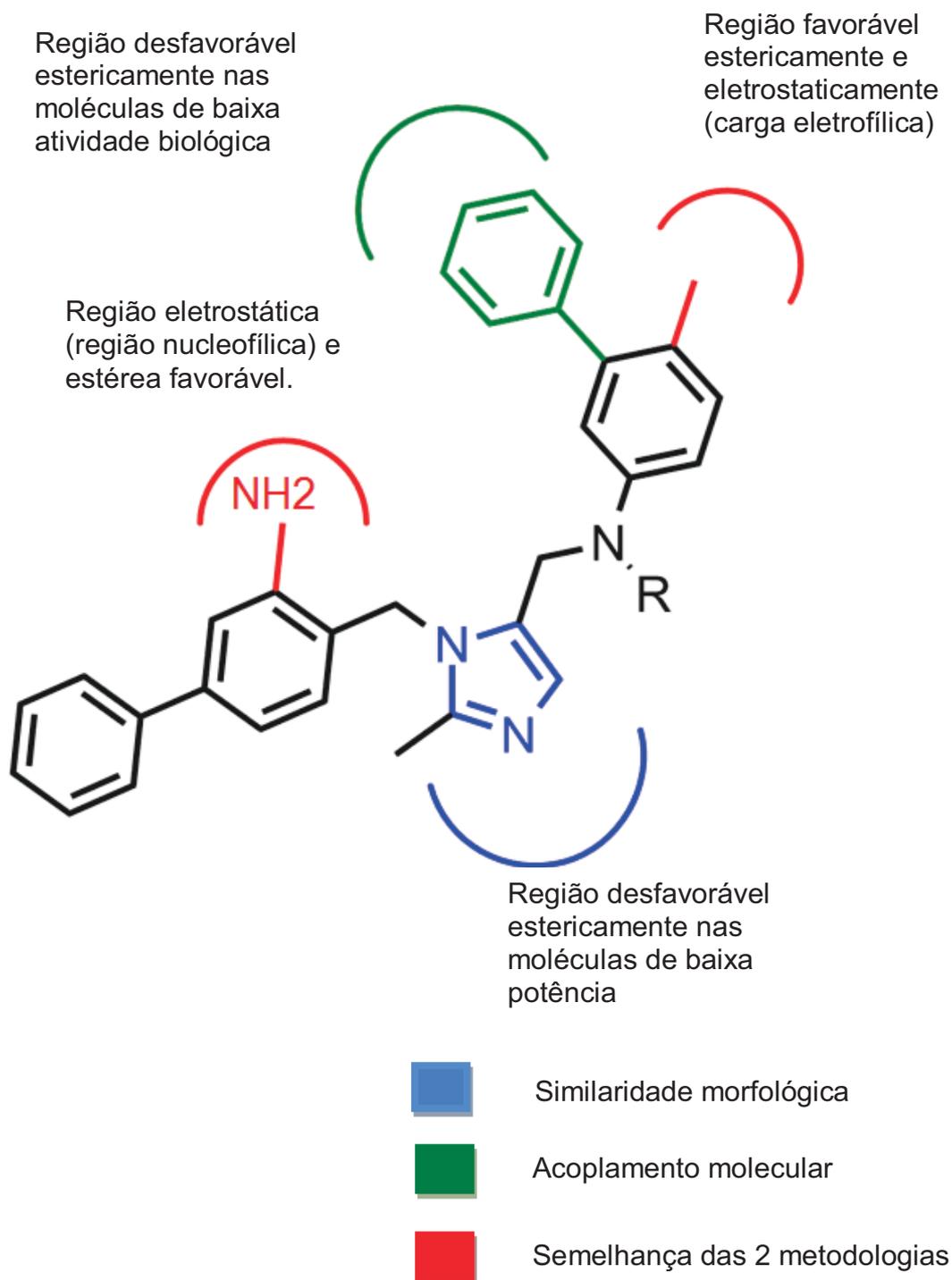
Os modelos de QSAR 2D e 3D realizados neste trabalho foram desenvolvidos com a preocupação de escolha de grupos treinamento e teste significativos para a análise, com diversidade química e de atividade, proporcionando satisfatória cobertura do espaço químico e biológico. O grupo treinamento teve influência para validação interna do modelo enquanto que o grupo teste foi de extrema importância para validação de moléculas que não foram utilizadas no modelo. A validação externa foi calculada por 2 fórmulas, uma que considera o valor preditivo somente do grupo teste e outra que inclui informações do grupo treinamento. O rigor estatístico considerado para o cálculo da validação externa é devido a importância deste para a análise do modelo.

Os modelos de QSAR 2D (HQSAR e QSAR baseado em descritores bidimensionais) apresentaram coeficientes de correlação e capacidade preditiva satisfatórios. A análise conjunta dos modelos, através do mapa de contribuição e descritores bidimensionais demonstrou que ramificação e índice de aromatização molecular desfavorece a atividade e a amina primária ligada ao anel aromático e informações estereoquímicas favorece a atividade.

Para complementar as informações oriundas da metodologia de QSAR 2D, foi realizado modelos de QSAR 3D. Os modelos de QSAR 3D derivados do alinhamento por acoplamento molecular com restrição do anel imidazol ao heme e similaridade morfológica sobrepostos ao alinhamento dos 5 compostos de maior atividade apresentaram bons coeficientes de correlação e capacidade preditiva satisfatórios. Esses modelos demonstraram a importância da amina primária ligada ao anel aromático, podendo interagir com a MET¹⁰⁶, o que está de acordo com os modelos de QSAR e HQSAR. Também substituintes eletrofílicos na posição *p*-cloro fenil são importantes para a atividade. Nas moléculas de baixa potência, o anel imidazol está em uma porção desfavorável estericamente, provavelmente devido ao fato do aumento da ramificação do esqueleto químico, demonstrado nos modelos de HQSAR e QSAR.

Um resumo do resultado das técnicas de QSAR 3D estão demonstradas abaixo. As informações das duas técnicas complementam as informações do

HQSAR e QSAR baseado em descritores topológicos, sendo importante, então, a análise conjunta das duas metodologias para esse conjunto de dados.



REFERÊNCIAS

- ABDAD-FRANCHA, F. et al. Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: A multi-scale appraisal of the tribe Rhodniini (Triatominae). **Acta Tropica**, v.110, n.2-3, p.159-177, 2009.
- ABLILASH, M. Docking methods. **International Journal of Pharma and Bio Sciences**, v.1, n.1, p.1-13, 2010.
- ACI, S. Molecular descriptor and similarity indices. In MARECHAL, E.; ROY, S.; LAFANECHERE, L. **Chemogenomics and Chemical Genetics**, New York: Springer. 2011, Cap.11, p.135-152.
- AKAMATSU, M. Current State and Perspectives of 3D-QSAR. **Current Topics in Medicinal Chemistry**, v.2, n.12, p.1381-1394, 2002.
- AMERY, W. K.; DE COSTER R.; CAERS, I. Ketoconazole from an antimycotic to a drug for prostate cancer. **Drug Dev.Res.**, v.8, n.1/4, p.299-307, 1986.
- ANDRADE, C.H. et al. Fragment-based and classical quantitative structure–activity relationships for a series of hydrazides as antituberculosis agents. **Molecular Diversity**, v.12, n.1, p.47-59, 2008.
- ANDRADE, S.G. et al. Biological, biochemical and molecular features of *Trypanosoma cruzi* strains isolated from patients infected through oral transmission during a 2005 outbreak in the state of Santa Catarina, Brazil: its correspondence with the new *T. cruzi* Taxonomy Consensus (2009). **Mem Inst Oswaldo Cruz**, v.106, n.8, p.948-956, 2011.
- ANDRICOPULO, A. D. et al. Structure-based drug design strategies in medicinal chemistry. **Curr.Top.Med. Chem.**, v.9, n.9, p.771-90, 2009.
- APT, W. et al. Treatment of chronic Chagas' disease with itraconazole and allopurinol. **The American journal of tropical medicine and hygiene**, v.59, n.1, p.133-8, 1998.
- APT, W. et al. Itraconazole or allupurinol in the treatment of chronic american trypanosomiasis the regression and prevention of electrocardiographic abnormalities suring 9 years of follow up. **Annals of tropical medicine and parasitology**, v.97, n.1, p.23-29, 2003.
- APT, B. W.; ZULANTAY, A. I. Estado actual en el tratamiento de la enfermedad de Chagas. **Rev Med Chile**, v.139, n.2, p.247-257, 2011.
- APT, W. Current and developing therapeutic agents in the treatment of Chagas disease. **Drug Design, Development and Therapy**, v.4, p.243 – 253, 2010.

- AVERY, M.A. et al. Structure–Activity Relationships of the Antimalarial Agent Artemisinin 6. The Development of Predictive In Vitro Potency Models Using CoMFA and HQSAR Methodologies. **J. Med. Chem.**, v.45, n.2, p.292–303, 2002.
- BASKIN, I.; VARNEK, A. Fragment Descriptors in SAR/ QSAR/QSPR Studies, Molecular Similarity Analysis and in Virtual Screening. In: VARNEK, A; TROPSHA, A. **Chemoinformatics Approaches to Virtual Screening**. 1ª ed. Royal Society of Chemistry: Cambridge, 2008. Cap. 1, p. 1-30.
- BASTOS, C.J.C. et al. Clinical Outcomes of Thirteen Patients with Acute Chagas Disease Acquired through Oral Transmission from Two Urban Outbreaks in Northeastern Brazil. **PLoS Neglected Tropical Disease**, v.4, n.6, p.1-6, 2010.
- BARBOSA-FERREIRA, J.M. et al. Cardiac Involvement in Acute Chagas' Disease Cases in the Amazon Region. **Arq. Bras. Cardiol.**, v.94, n.6, e83-e85, 2010.
- BEACH, D.H.; GOAD, L. J.; HOLTZ G. G. Effects of ketoconazole on sterol biosynthesis by *Trypanosoma cruzi* epimastigotes. **Biochem. Biophys. Res. Commun.**, v.136, n.3, p. 851-856, 1986.
- BELTRÃO, H.B.M. et al. Investigation of two outbreaks of suspected oral transmission of acute Chagas disease in the Amazon region, Para State, Brazil, in 2007. **Trop Doct.**, v.39, n.4, p.231-2, 2009.
- BENVENUTI, L.A. et al. Chronic American trypanosomiasis: parasite persistence in endomyocardial biopsies is associated with high-grade myocarditis. **Ann. Trop. Med. Parasitol.**, v.102, n.6, p.481–487, 2008.
- BEYRER, C. et al. Neglected diseases, civil conflicts, and the right to health. **Lancet**, v.370, n.9587, p.619-27, 2007.
- BRODNIEWICZ, T.; GRYNKIEWICZ, G. Preclinical drug development. **Acta Poloniae Pharmaceutica - Drug Research**, v.67, n.6, p.579-586, 2010.
- BROWN, D. Unfinished business: target-based drug discovery. **Drug Discovery Today**, v.12, n.23/24, p. 1007-1012, 2007.
- BUCKNER, F. et al. A class of sterol 14-demethylase inhibitors as anti-*Trypanosoma cruzi* agents. **PNAS**, v.100, n.25, p.15149–15153, 2003.
- BURCHAN, N.S. et al. The role of translational bioinformatics in drug discovery. **Drug Discovery Today**, v.16, n.9/10, p.426-434, 2011.
- CHAGAS, C., Nova tripanosomiase humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen, n. sp., agente etiológico do nova entidade mórbida do homem. **Mem. Inst. Oswaldo Cruz. Mem. Inst. Oswaldo Cruz** , v.1, p. 159–218, 1909.

CHATELAIN, E.; IOSET, J.R. Drug discovery and development for neglected diseases: the DNDi model. **Drug Des Devel Ther.** , v.5, p.175–181, 2011.

CHAUDHAERY, S.S.; ROY, K.K; SAXENA, A.K.Consensus Superiority of the Pharmacophore-Based Alignment, Over Maximum Common Substructure (MCS): 3D-QSAR Studies on Carbamates as Acetylcholinesterase Inhibitors. **J. Chem. Inf. Model.**, v.49, n.6, p.1590–1601, 2009.

CHE, X. et al. New azoles with potent antifungal activity: design, synthesis and molecular docking. **Eur J Med Chem.**, v.44, n.10, p.4218-26, 2009.

CHEN, C. K. et al. Structural Characterization of CYP51 from *Trypanosoma cruzi* and *Trypanosoma brucei* Bound to the Antifungal Drugs Posaconazole and Fluconazole. **PLoS Negl Trop Dis**, v.4, n.4, e651, 2010.

CHEN, H.et al. On evaluating molecular-docking methods for pose prediction and enrichment factors. **J. Chem. Inf. Model**, v.46, n.1, p.401-415, 2006.

CHEN, D. et al. Holographic QSAR of selected esters. **Chemosphere**, v.57, n.11, p.1739–1745, 2004.

CHIRAC, P.; TORREELE, E. Global framework on essential health R&D. **TheLancet**, v.367, n.9522, p.1560-1561, 2006.

CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. **J. Chem. Inf. Model.**, v.51, n.9, p. 2320–2335, 2011.

CLAYTON, J. Chagas disease: pushing through the pipeline. **Nature**, v.465, n.7301, p.S12-S15, 2010.

CLEVES, A.E; JAIN, A.N. Robust Ligand-Based Modeling of the Biological Targets of Known Drugs. **J. Med. Chem.**, v.49, n.10, p.2921–2938, 2006.

CLEVES, A.E.; JAIN, A.N. Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. **Journal of Computer-Aided Molecular Design**, v.22, n.3-4, p.147-159, 2008.

COHEN, J.; DIBNER, M.S.; WILSON, A. Development of and Access to Products for Neglected Diseases. **PLoS ONE**, v.5, n.5, e10610, 2010.

CONSONNI, V.; BALLABIO, D.; TODESCHINI, R. Comments on the Definition of the Q^2 Parameter for QSAR Validation. **J. Chem. Inf. Model.**, v.49, n.7, p.1669-1678, 2009.

CORDOVA, E. et al. Reactivation of Chagas disease with central nervous system involvement in HIV-infected patients in Argentina, 1992–2007. **Int. J. Infect. Dis.**, v.12, n. 6, p.587–592, 2008.

CORRALES, M. et al. Comparative efficacies of TAK-187, a longlasting ergosterol biosynthesis inhibitor, and benznidazole in preventing cardiac damage in a murine model of Chagas' disease. *Antimicrob. Agents Chemother.*, v.49, n.4, p.1556–1560, 2005.

COURA, J.R. Chagas disease: what is known and what is needed – A background article. *Mem Inst Oswaldo Cruz*, v.102, p.113-122, 2007.

COURA, J. R.; VIÑAS, P. A. Chagas disease: a new worldwide challenge. *Nature*, v.24, n.7302, S6-S7, 2010.

CRAMER, R.D., III; PATTERSON, D.E.; BUNCE, J.D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.*, v.110, p.5959-5967, 1988.

CROFT, S.L. et al. Chemotherapy of trypanosomiases and leishmaniasis. *Trends Parasitol.*, v.21, n.11, p.508–512, 2005.

CUISSART, F. et al. The Maximum Common Substructure as a Molecular Depiction in a Supervised Classification Context: Experiments in Quantitative Structure/Biodegradability Relationships. *J. Chem. Inf. Comput. Sci.*, v.42, n.5, p.1043–1052, 2002.

CUNHA-NETO, E.; KALIL, J. Autoimmunity in Chagas' heart disease. *São Paulo Medicinal Journal*, v.113, n.3, p.757-766, 1995.

DA CHEN et al. Holographic QSAR of selected esters. *Chemosphere*, v.57, n.11, p.1739-1745, 2004.

DIAS, J.C.P; SILVEIRA, A.C; SCHOFIELD, C.J. The Impact of Chagas Disease Control in Latin America - A Review. *Mem. Inst. Oswaldo Cruz*, v.97, n.5, p.603-612, 2002.

DIAS, R.; AZEVEDO, W.F. Molecular Docking Algorithms. *Current Drug Targets*, v.9, p.1040-1047, 2008.

DIAS, J.P. et al. Acute Chagas disease outbreak associated with oral transmission. *Revista da Sociedade Brasileira de Medicina Tropical*, v.41, n.3, p.296-300, 2008.

DIEZ, M. et al. Usefulness of PCR strategies for early diagnosis of Chagas' disease reactivation and treatment follow-up in heart transplantation. *Am. J. Transplant.*, v.7, n.6, p. 1633–1640, 2007.

DNDI. Doenças Negligenciadas. Doença de Chagas. **DNDI**, 2010. Disponível em <<http://www.dndi.org.br/pt/doencas-negligenciadas/doenca-de-chagas.html>>. Acesso em 20 de jul. de 2011.

D'OCA, G. Computational medicinal chemistry: part II. **Future Med. Chem.**, v.3, n.6, p.641–642, 2011.

DRIE, J.H.V. Monty Kier and the Origin of the Pharmacophore Concept. **Internet Electronic Journal of Molecular Design**, v.6, p. 271–279, 2007.

DUDEK, S. Z.; ARODZ, T.; GÁLVEZC, J. Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. **Combinatorial Chemistry & High Throughput Screening**, v.9, p.213-228, 2006.

EASTMAN, R.T. et al. Thematic review series: lipid postranslational modifications. Fighting parasitic disease by blocking protein farnesylation. **Journal of Lipid Research**, v.47, n.2, p. 233-240, 2006.

ELUMALAI, P. et al. Pharmacophore modeling, virtual screening and docking studies to identify novel HNMT inhibitors. **Journal of the Taiwan Institute of Chemical Engineers**, v.43, n.4, p.493-503, 2012.

EPTING, C. L.; COATES, B. M., ENGMAN, D. M. Molecular mechanisms of host cell invasion by *Trypanosoma cruzi*. **Experimental Parasitology**, v.126, p.283–291, 2010.

FERREIRA, R.S.; OLIVA, G.; ANDRICOPULO, A.D. Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos. **Quím.Nova**, v.34, n.10, 2011.

FERREIRA, M.M.C.; MONTANARI, C.A.; GAUDIO, A.C. Seleção de variáveis em QSAR. **Quim. Nova**, v.25, n.3, p.439-448, 2002.

FIORELLI, A.I. et al. Later evolution after cardiac transplantation in Chagas' disease. **Transplant.Proc.**, v.37, n.6, p.2793–2798, 2005.

FLOHÉ, L. The trypanothione system and the opportunities it offers to create drugs for the neglected kinetoplast diseases. **Biotechnology Advances**, v.30, n.1, p.294–301, 2012.

GARCIA, S. et al. Treatment with benznidazole during the chronic phase of experimental Chagas' disease decreases cardiac alterations. **Antimicrob. Agents Chemother.**, v.49, n.4, p.1521–1528, 2005.

GARCIA, T. S.; HONÓRIO, K. M. Two-Dimensional Quantitative Structure-Activity Relationship Studies on Bioactive Ligands of Peroxisome Proliferator-Activated Receptor δ . **J. Braz. Chem. Soc.**, v.22, n.1, p.65-72, 2011.

GARG, N.; TARLETON, R.L. Genetic immunization elicits antigen-specific protective immune responses and decreases disease severity in *Trypanosoma cruzi* infection. **Infect. Immun.**, v.70, n.10, p.5547–5555, 2002.

GELB, M.H. Drug Discovery for Malaria: A Very Challenging and Timely Endeavor.

Curr Opin Chem Biol., v.11, n.4, p. 440–445, 2007.

GIACOMINO, A. et al. The role of chemometrics in single and sequential extraction assays: A Review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques. **Analytica Chimica Acta**, v.688, n.2, p.122–139, 2011.

GOHDAA, K. et al. A CoMFA analysis with conformational propensity: An attempt to analyze the SAR of a set of molecules with different conformational flexibility using a 3D-QSAR method. **Journal of Computer-Aided Molecular Design**, v.14, p.265–275, 2000.

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

GUBBINS, P.O; ANAISSIE, E.J. Antifungal therapy. In ANAISSIE, E.J; MCGINNIS, M.R; PFALLER, M.A. **Clinical Mycology**. 2^a ed. China: Elsevier, 2009. Cap 7, p.161-196.

GUIDO, R. V. C.; OLIVA, G. Structure-Based Drug Discovery for Tropical Disease. **Current Topics in Medicinal Chemistry**, v.9, p.824-843, 2009.

GUIDO, R.V.C et al. Structural Basis for Selective Inhibition of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase: Molecular Docking and 3D QSAR Studies. **J. Chem. Inf. Model.**, v.48, p.918–929, 2008.

GUIDO, R.V.C et al. Classical and Hologram QSAR Studies on a Series of Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase. **QSAR Comb. Sci.** v.27, n.6, p.768 – 781, 2008.

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, 2010.

HAWKINS, P.C.D.; SKILLMAN, A.G.; NICHOLLS, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. **J. Med. Chem**, v.50, n.1, p.74-82, 2007.

HARTMANN, M. A. Plant sterols and the membrane environment. **Trends in Plant Science**, v.3, n.5, p.170-175, 1998.

HASSLOCHER-MORENO, A.M. et al. Safety of benznidazole use in the treatment of chronic Chagas' disease. **J. Antimicrob. Chemother.**, v.67, n.5, p.1261-1266, 2012.

HOELDER, S.; CLARKE, P.A.; WORKMAN, P. Discovery of small molecule cancer drugs: Successes, challenges and opportunities. **Molecular Oncology**, v.6, n.2, p.155-176, 2012.

HONÓRIO, K.M. et al. 3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators. **Journal of Molecular Graphics and Modelling**, v.25, n.6, p. 921–927, 2007.

HOTEZ, P.J. et al. The Neglected Tropical Diseases of Latin America and the Caribbean: A Review of Disease Burden and Distribution and a Roadmap for Control and Elimination. **PLoS Negl Trop Dis**, v.2, n.9, e300, 2008.

HOTEZ, P.J. et al. Incorporating a Rapid-Impact Package for Neglected Tropical Diseases with Programs for HIV/AIDS, Tuberculosis, and Malaria. **PLoS Med**, v. 3, n.5, e102, 2006.

HUGHES, J.P. et al. Principles of early drug discovery. **British Journal of Pharmacology**, v.162, n.6, p.1239-1249, 2011.

ISHIDA, K. et al. Two squalene synthase inhibitors, E5700 and ER-119884, interfere with cellular proliferation and induce ultrastructural and lipid profile alterations in a *Candida tropicalis* strain resistant to fluconazole, itraconazole, and amphotericin B. **J. Infect. Chemother.**, v.17, n.4, p.563–570, 2011.

JAIN, A. N. Ligand-Based Structural Hypotheses for Virtual Screening. **J. Med. Chem.**, v.47, n.4, p.947-961, 2004.

JAIN, A. N. Morphological similarity: A 3D molecular similarity method correlated with protein-ligand recognition. **Journal of Computer-Aided Molecular Design**, v.14, n.2, p.199–213, 2000.

JÓNSDÓTTIR, S. O., FLEMMING, S. J.; BRUNAK, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. **Bioinformatics**, v.21, n.10, p.2145-2160, 2005.

JORGENSEN, W. L. Progress and issues for computationally guided lead discovery and optimization. In: MERZ, K. M.; RINGE, D.; REYNOLDS, C.H. **Drug design: structure- and ligand-based approaches**. New York: Cambridge, 2010. Cap 1, p.1-16.

KIRTON, S. et al. Prediction of Binding Modes for Ligands in the Cytochromes P450 and Other Heme-Containing Proteins. **PROTEINS: Structure, Function, and Bioinformatics**, v.58, n.4, p.836–844, 2005.

KITCHEN, D. B et al. Docking and scoring in virtual screening for drug discovery: methods and application. **Nature Reviews Drug Discovery**, v.3, n.11, p.935-949, 2004.

KOEPPEN, H. et al. Ligand-Based Virtual Screening. In: MANNHOLD, R.; KUBINYI, H.; FOLKERS, G. **Virtual Screening: Principles, Challenges, and Practical Guidelines**. Weinheim: WILEY-VCH, 2011, cap. 3.

KUBINYI, H. Comparative Molecular Field Analysis (CoMFA). In: **The Encyclopedia of Computational Chemistry**. John Wiley Sons, 1998, p. 448-460.

KUBINYI, H. QSAR E 3D QSAR in drug designer. Part 1: methodology. **DDT**, v.2, n.11, p.457-466, 1997.

KUBINYI, H. QSAR and 3D QSAR in drug design. Part 2: applications and problems. **DDT**, v.2, n.12, p. 457-467, 1997.

LEACH, A. R.; GILLET, V. J. Computational Models. In _____: **An introduction to chemoinformatics**. Netherland: Springer, 2007. Cap. 4. p. 75-97.

LEE, C.H.; HUANG, H.C.; JUAN, H.F. Reviewing Ligand-Based Rational Drug Design: The Search for an ATP Synthase Inhibitor. **International Journal of Molecular Sciences**, v.12, n.8, p.5304-5318, 2011.

LEONARD, J.T; ROY,K. On Selection of Training and Test Sets for the Development of Predictive QSAR models. **QSAR Comb. Sci.**, v.25, n.3, p.235 – 251, 2006.

LEPESHEVA, G.I. et al. CYP51 from *Trypanosoma cruzi*: A phyla-specific residue in the B' helix defines substrate preferences of sterol 14 α -demethylase. **Journal of Biological Chemistry**, v.281, n.6, p. 3577-3585, 2006.

LEPESHEVA, G.I. et al. Sterol 14 α -Demethylase as a Potential Target for Antitrypanosomal Therapy: Enzyme Inhibition and Parasite Cell Growth. **Chemistry & Biology**, v.14, n.11, p.1283-1293, 2007.

LEPESHEVA, G.I. et al. Structural Insights into Inhibition of Sterol 14 α -Demethylase in the Human Pathogen *Trypanosoma cruzi*. **The Journal of Biological Chemistry**, v.285, n.33, p.25582–25590, 2010.

LI, Y.Y.; AN, J.; JONES, S.J.M.A Computational Approach to Finding Novel Targets for Existing Drugs. **Drugs. PLoS Comput Biol**, v.7, n.9, e1002139, 2011.

LIEBESCHUETZ, J.W.; COLE, J.C.; KORB, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. **J. Comput. Aided Mol. Des.**, v.26, n.6, p.737-748, 2012.

LOMBARDINO, J. G.; LOWE, J. A. 3rd. The role of the medicinal chemist in drug discovery – then and now. **Nat. Rev. Drug Discov.**, v.3, n.10, p.853-62, 2004.

LOWIS, D. HQSAR: A new, highly predictive QSAR technique. **Tripos Technical Notes**, v.1, n.5. p.1-10, 1997.

MAGALHÃES, C.S.; BARBOSA, H.J.C; DARDENNE, L.E. Métodos de docking receptor-ligante para o desenho racional de compostos bioativos. In: MORGON, N.H; COUTINHO, K. **Métodos de Química Teórica e Modelagem Molecular**. São Paulo: Editora Livraria da Física, 2007, Cap. 14, p. 489-532.

- MALDONADO, R.A. et al. Experimental chemotherapy with combinations of ergosterol biosynthesis inhibitors in murine models of Chagas' disease. **Antimicrob. Agents Chemother.**, v.37, p.1353–135, 1993.
- MADHAVAN, T. et al. Various atomic charge calculation schemes of CoMFA on HIF-1 inhibitors of moracin analogs. **International Journal of Quantum Chemistry**, v.112, n.4, p.995–1005, 2012.
- MARAN, U. et al. Molecular descriptors from two-dimensional chemical structure. In: CRONIN, T.D.M; MADDEN, J.C. **In silico toxicology: Principles and applications**. Royal Society: Ingleterra, 2010, cap. 6, p. 148-189.
- MARIN-NETO, J.A. et al. Rationale and design of a randomized placebo-controlled trial assessing the effects of etiologic treatment in Chagas' cardiomyopathy: the benznidazole evaluation for interrupting trypanosomiasis (BENEFIT). **Am. Heart J.**, v.156, n.1, p.37–43, 2008.
- MASSAD, E. The elimination of Chagas' disease from Brazil. **Epidemiol. Infect.**, v.136, n.9, p.1153–1164, 2008.
- MAURI, A. et al. DRAGON software: an easy approach to molecular descriptor calculations. **Match**, v.56, n.2, p.237-248, 2006.
- MAYA, J.D. et al. Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. **Comparative Biochemistry and Physiology**, v.146, n.4, p. 601–620, 2007.
- MAYA, J.D. et al. Chagas disease: Present status of pathogenic mechanisms and chemotherapy. **Biological Research**, v.43, n.3, p.323-331, 2010.
- MAYR, L.M.; BOJANIC, D. Novel trends in high-throughput screening. **Current Opinion in Pharmacology**, v.9, n.5, p.580–588, 2009.
- MEANWELL, N. A. Improving Drug Candidates by Design: A Focus on Physicochemical Properties As a Means of Improving Compound Disposition and Safety. **Chem. Res. Toxicol.**, v.24, n.9, p.1420–1456, 2011.
- MEEK, P.; MOYNA, G.; ZAUHAR, R. Computer techniques: identifying similarities between small molecules. In: SHAYNE COX GAD, S. C. **Preclinical Development Handbook: ADME and Biopharmaceutical Properties**. 1^a ed., New Jersey: John Wiley & Sons, Inc. Hoboken, 2008. Cap. 2, p. 47- 86.
- MEJIA, A.M. et al. Benznidazole-Resistance in *Trypanosoma cruzi* Is a Readily Acquired Trait That Can Arise Independently in a Single Population. **J. Infect. Dis.**, v.206, n.2, p.220-8, 2012.
- MITTAL, R.R. et al. Partial Charge Calculation Method Affects CoMFA QSAR Prediction Accuracy. **J. Chem. Inf. Model.**, v.49, n.3, p.704–709, 2009.

- MITTAL, R. R.; A. MCKINNON, R. A.; SORICH, M.J. The Effect of Molecular Fields, Lattice Spacing and Analysis Options on CoMFA Predictive Ability. **QSAR Comb.Sci.**, v.28, n.6-7, p.637 – 644, 2009.
- MOLYNEUX, D.H.; HOTEZ, P.J.; FENWICK, A. “Rapid-Impact Interventions”: How a Policy of Integrated Control for Africa's Neglected Tropical Diseases Could Benefit the Poor. **PLoS Med.**, v.2, n.11, e336, 2005.
- MORAN, M. et al. Neglected Disease Research and Development: How Much Are We Really Spending? **PLoS Medicine**, v.6, n.2, e1000030, 2009.
- MOTA, S.G; BARROS, T.F.; CASTILHO, M.S. 2D QSAR Studies on a Series of Bifonazole Derivatives with Antifungal Activity. **J. Braz. Chem. Soc.**, v.20, n.3, p.451-459, 2009.
- MOURITSENA, O. G.; ZUCKERMANN, M. J. What’s So Special About Cholesterol? **Lipids**, v.39, n.11, p.1101-1113, 2004.
- NÓBREGA, A.A. et al. Oral Transmission of Chagas Disease by Consumption of Açaí Palm Fruit, Brazil. **Emerging Infectious Diseases**, v.15, n.4, p.653-655, 2009.
- NWAKA, S.; HUDSON, A. Innovative lead discovery strategies for tropical diseases. **Nature Reviews Drug Discovery**, v.5, n.11, p.941-955, 2006.
- NWAKA, S.; RIDLEY, R. G. Virtual drug discovery and development for neglected disease through public-private partnerships. **Nature Reviews Drug Discovery**, v.2, n.11, p. 919-928, 2003.
- ODDS, F. C.; BROWN, A. J. P.; GOW, N. A. R. Antifungal agents: mechanisms of action. **TRENDS in Microbiology**, v.11, n.6, p.272-279, 2003.
- OJHA, P.K; ROY, K. Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection. **Chemometrics and Intelligent Laboratory Systems**, v.109, n.2, p.146–161, 2011.
- OLIVEIRA, M. Neurologic Manifestations of Chagas Disease. **Current Neurology and Neuroscience Reports**, v.11, n.6, p.536-542, 2011.
- ORDOG, R.; GROLMUSZ, V. **Evaluating Genetic Algorithms in Protein-Ligand Docking**. In MANDOIU, I.; SUNDERRAMAN, R. ZELIKOVSLY, A. Bioinformatics research and applications, 2008
- OECD. Guidance Document on the Validation of (Q)SAR Models. Paris, 2007. Disponível em <http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono%282007%292>. Acesso em 23 de Março de 2012.

- OTTO, M. Pattern recognition and classification. In_____. **Chemometrics: Statistics and Computer Application in Analytical Chemistry**. Weinhein, New York, Chichester, Brisbane, Singapore, Toronto: WILEY-VCH, 2007, Cap. 5, p.121-181.
- PADHY, B.M.; GUPTA, Y.K. Drug repositioning: Re-investigating existing drugs for new therapeutic indications. **Technology Review**, v.57, n.2, p.153-160, 2011.
- PEREZ-FUENTES, R. et al. Severity of chronic Chagas disease is associated with cytokine/antioxidant imbalance in chronically infected individuals. **Int. J. Parasitol.**, v.33, n.3, p.293–299, 2003.
- PEREZ-MOLINA, J. A. Use of benznidazole to treat chronic Chagas' disease: a systematic review with a meta-analysis. **Journal of Antimicrobial Chemotherapy**, v.64, n.6, p.1139-1147, 2009.
- PETERSON, S.D.; SCHAAL, W.; KARLÉN, A. Improved CoMFA Modeling by Optimization of Settings. **J. Chem. Inf. Model.**, v.46, n.1, p.355-364, 2004.
- PINK, R. et al. Opportunities and challenges in antiparasitic drug discovery. **Nature Reviews Drug Discovery**, v.4, n.9, p.727-740, 2005.
- PINTO, A. Y. N. et al. Urban outbreak of acute Chagas disease in Amazon region of Brazil: four-year follow-up after treatment with benznidazole. **Rev Panam Salud Publica/Pan Am J Public Health**, v.25, n.1, p.77-83, 2009.
- PODUST, L.M.; POULOS, T.L.; WATERMAN, M.R. Crystal structure of cytochrome P450 14a-sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. **Proc Natl Acad Sci U S A**, v.98, n.6, p.3068-73, 2001.
- PONTES, V. M. O. et al. Reações adversas em pacientes com doença de Chagas tratados com benzonidazol, no Estado do Ceará. **Rev. Soc. Bras. Med. Trop.**, v. 43, n.2, p. 182-187, 2010.
- RASSI JR, A., RASSI, A; MARIN-NETO, J.A. Chagas disease. **The Lancet**, v.375, n.9723, p.1388-1402, 2010.
- RAHNASTO, M.K. et al. Identification of novel CYP2A6 inhibitors by virtual screening. **Bioorganic & Medicinal Chemistry**, v.19, n.23, p.7186–7193, 2011.
- ROTH, B.D. The discovery and development of atorvastatin, a potent novel hypolipidemic agent. **Prog Med Chem.**, v.40, p.1-22, 2002.
- ROY, P.P; ROY, K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. **QSAR Comb. Sci.**, v.27, n.3, p.302 – 313, 2008.
- SALUM, L. B.; ANDRICOPULO, A. D. Fragment-based QSAR: perspectives in drug design. **Mol Divers**, v.13, n.3, p.277–285, 2009.

- SAMS-DODD, F. Drug discovery: selecting the optimal approach. **Drug Discovery Today**, v.11, n. 9/10, p.465-472, 2006.
- SCHEIDER, G. Computational medicinal chemistry. **Future Med. Chem.**, v.3, n 4, p.393–394, 2011.
- SCHIJMAN, A.G. et al. Trypanosoma cruzi DNA in cardiac lesions of Argentinean patients with end-stage chronic Chagas heart disease. **Am. J. Trop.Med. Hyg.**, v.70 n.2, p.210–220, 2004.
- SCHUURMANN, G. External validation and prediction employing the predictive squared correlation coefficient-test set activity mean vs training set activity mean. **J. Chem. Inf. Model.**, v.48, n.11, p.2140-2145, 2008.
- SCIOR, T.et al. How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review. **Current Medicinal Chemistry**, v.16, n.32, p.4297-4313, 2009.
- SCHMUNIS, G.A; YADON, Z.E. Chagas disease: a Latin American health problem becoming a world health problem. **Acta Trop.**, v.115, n.1-2, p.14-21, 2010.
- SEEL, M.; TURNER, D.B; WILLETT, P. Effect of Parameter Variations on the Effectiveness of HQSAR Analyses. **Quant. Struct.-Act. Relat.**, v.18, n.3, p.245-252, 1999.
- SHEN, J. et al. Estimation of ADME Properties with Substructure Pattern Recognition. **J. Chem. Inf. Model.**, v.50, n.6, p.1034–1041, 2010.
- SHIKANAI-YASUDA, M.A.; CARVALHO, N.B. Oral Transmission of Chagas Disease. **Clin Infect Dis.**, v.54, n.6, p.845-852, 2012.
- SILVA, C.F. et al. In vitro and in vivo investigation of the efficacy of arylimidamide DB1831 and its mesylated salt form - DB1965 - against Trypanosoma cruzi infection. **PLoS ONE**, v.7, n.1, e30356, 2012.
- SILVA, F. C. et al. Chagas Disease: Challenges in Developing New Trypanocidal Lead Compounds. **Rev. Virtual Quim.**, v.4, n.1, p.46-72, 2012.
- SLINKER, B.K.; GLANTZ, S.A. Multiple Linear Regression Accounting for Multiple Simultaneous Determinants of a Continuous Dependent Variable. **Circulation**, v.117, n.13, p.1732-1737, 2008.
- SOUSA, S.F.; FERNANDES, P.A.; RAMOS, M.J. Protein–ligand docking: Current status and future challenges. **Proteins: Structure, Function, and Bioinformatics**, v.65, n.1, p.15–26, 2006.
- SPANOVA, M.; DAUM, G. Squalene – biochemistry, molecular biology, process biotechnology, and applications. **Eur. J. Lipid Sci. Technol.**, v.113, n.11, p.1299–1320, 2011.

SURYADEVARA, P. K. et al. Structurally Simple Inhibitors of Lanosterol 14 α -Demethylase Are Efficacious In a Rodent Model of Acute Chagas Disease. **J. Med Chem**, v.52, n.12, p.3703-3715, 2009.

SWINNEY, D.C.; ANTHONY, J. How were new medicines discovered? **Nature Reviews Drug Discovery**, v.10, n.7, p.507-519, 2011.

TARLETON, R.L. Chagas disease: a role for autoimmunity? **Trends Parasitol.**, v.19, n.10, p.447-451, 2003.

TASHIRO, E.; IMOTO, M. Target identification of bioactive compounds. **Bioorganic & Medicinal Chemistry**, v.20, n.6, p.1902-1909, 2012.

TEOFILO, R.F.; MARTINSA, J.P.A; FERREIRA, M.C. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. **J. Chemometrics**, v.23, p.32-48, 2009.

TITOV, D.V.; LIU, J.O. Identification and validation of protein targets of bioactive small molecules. **Bioorg. Med. Chem.**, v.20, n.6, p.1922-1928, 2012.

TODESCHINI, R. et al. Mobydigs: Software for regression and classification models by genetics algorithms. **Data Handling in Science and Technology**, v.23, p.141-167, 2003.

TODESCHINI, R.; CONSONNI, V. **Handbook of Molecular Descriptors**. Weinheim; Germany: Wiley-VCH, 2000.

TODESCHINI, R.; CONSONNI, V.; PAVAN, M. MobyDigs: Software for Regression and Classification Models by Genetic Algorithms. In: LEARD, R. **Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks** Elsevier: Amsterdam, 2003, cap. 5, p. 141-167.

TONG, W. et al. Evaluation of Quantitative Structure-Activity Relationship Methods for Large-Scale Prediction of Chemicals Binding to the Estrogen Receptor. **J. Chem. Inf. Comput. Sci.**, v.38, n.4, p.669-677, 1998.

URBINA, J.A. Chemotherapy of Chagas Disease. **Curr Pharm Design**, v.8, n.4, p.287-295, 2002.

URBINA, J. A.; DOCAMPO, R.. Specific chemotherapy of Chagas disease: controversies and advances. **Trends Parasitol**, v.19, n.11, p.495-501, 2003.

URBINA, J. A. Ergosterol biosynthesis and drug development for Chagas disease. **Mem Inst Oswaldo Cruz**, v.104, n.1, p.311-318, 2009.

URBINA, J.A. Specific chemotherapy of Chagas disease: Relevance, current limitations and new approaches. **Acta Tropica**, v.115, n.1-2, p.55-68, 2010.

- VAGO, A.R. et al. Genetic characterization of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. **Am. J. Pathol.**, v.156, n.5, p.1805–1809, 2000.
- VAIDIAN, A.K., et al. Chagas' disease and AIDS. **Kinetoplastid Biol. Dis.**, v.3, n.1, p.1-6, 2004.
- VERLI, H.; BARREIRO, E.J.. Um paradigma da química medicinal: a flexibilidade dos ligantes e receptores. **Quim.Nova**, v.28, n.1, p.95-102, 2005.
- VILLELA, M.M. et al. Avaliação do Programa de Controle da Doença de Chagas em relação à presença de *Panstrongylus megistus* na região centro-oeste do Estado de Minas Gerais, Brasil. **Cad. Saúde Pública, Rio de Janeiro**, v.25, n.4, p.907-917, 2009.
- VIRTANEN, S.I.; PENTIKAINEN, T. Efficient Virtual Screening Using Multiple Protein Conformations Described as Negative Images of the Ligand-Binding Site. **J. Chem. Inf. Model.**, v.50, n.6, p.1005–1011, 2010.
- WANG, X. et al. Molecular hologram derived quantitative structure–property relationships to predict physico-chemical properties of polychlorinated biphenyls. **Chemosphere**, v.51, n.7, p.617–632, 2003.
- WANG, G. et al. Understanding the Aquatic Toxicity of Pesticide: Structure-Activity Relationship and Molecular Descriptors to Distinguish the Ratings of Toxicity. **QSAR Comb. Sci.**, v.28, n.11-12, p.1418 –1431, 2009.
- WILKINSON, S.R. et al. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. **Proc. Natl. Acad. Sci.**, n.105, n.13, p.5022-7, 2008.
- WILSON, R.M; DANISHEFSKY, S.J. Small Molecule Natural Products in the Discovery of Therapeutic Agents: The Synthesis Connection. **J. Org. Chem.**, v.71, n.22, p. 8329–8351, 2006.
- WOLD,S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v.2, n.1-3, p.37-52, 1987.
- WORLD HEALTH ORGANIZATION. **First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases**. France: WHO, 2010.
- YAMAGATA, Y.; NAKAGAWA, J. Control of Chagas Disease. **Advances in Parasitology**, v.61, p.129-165, 2006.
- YANG, S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug Discovery Today**, v.15, n.11/12, p.444-450, 2010.
- YUAN, S. et al. Identification of maximal common substructures in structure/activity studies. **Analytica Chimica Acta**, v.235, p. 239-241, 1990.

ZHANG, L.; TARLETON, R.L.. Parasite persistence correlates with disease severity and localization in chronic Chagas' disease. **J. Infect. Dis.**, v.180, n.2, p.480–486, 1999.

ZHANG, L. et al. How to generate reliable and predictive CoMFA models. **Curr Med Chem.**, v.18, n.6, p.923-30, 2011.

ZHENG, J. et al. Three-dimensional quantitative structure-activity relationships of pyrrolopyridinone as cell division cycle kinase inhibitors by CoMFA and CoMSIA. **J Mol Model.**, v.17, p.2113–2130, 2011.

ZHUO. Y. et al. Three-dimensional QSAR analyses of 1,3,4-trisubstituted pyrrolidine-based CCR5 receptor inhibitors. **European Journal of Medicinal Chemistry**, v.43, n.12, p.2724-2734, 2008.